NAT'L INST. OF STAND & TECH R.I.C.
A11103 994436

NIST
PUBLICATIONS
REFERENCE

NBSIR 78-1446

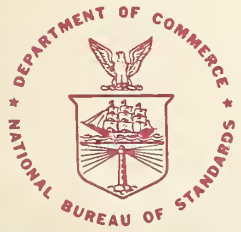
Management of Data Elements in Information Processing Proceedings of the Third National Symposium, Held at NBS on 1977 September 28-30

Sponsored by the National Bureau of Standards and the
American National Standards Institute Committee X3L8

Institute for Computer Sciences & Technology
National Bureau of Standards
Washington, D.C. 20234

April 1978

Final



U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

QC
100
.U56
no.78-1446
1978

DATE DUE

GAYLORD

NATIONAL BUREAU
OF STANDARDS
LIBRARY
AUG 10 1978 -NE
781833
QC100
.U56
78-1446
1977

NBSIR 78-1446

MANAGEMENT OF DATA ELEMENTS IN INFORMATION PROCESSING; PROCEEDINGS OF THE THIRD NATIONAL SYMPOSIUM, HELD AT NBS ON 1977 SEPTEMBER 28-30

Hazel E. McEwen, Editor

Institute for Computer Sciences & Technology
National Bureau of Standards
Washington, D.C. 20234

April 1978

Final

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, *Secretary*

Dr. Sidney Harman, *Under Secretary*

Jordan J. Baruch, *Assistant Secretary for Science and Technology*

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

Table of Contents[1]

Page

Energy Information - Top Priority for Today 1
Ted M. Albert

Standardization of Data Elements and Machine-Readable Symbol to Facilitate Blood Banking 7
Eric Brodheim, Sc.D.

A Technical Profile of Representative Government Developed Data Element Dictionary/Directory Systems 19
Paul K. Brown

Cargo Data Interchange System (CARDIS). 25
Robert J. Cavanaugh

American National Standards Institute (ANSI) 35
Ward Duel, MPh

Management of Data Elements in Information Processing (Needs for Standards in the Health Care Field and a Sampling of Current Applications) 39
Gerald J. Duffy

The Need for Standards in Data Transmission 43
Peter J. Finnerty, Esq.

Data Processing in Natural History Museums 47
T. Gary Gautier, Stanwyn G. Shetler and Craig Black

Commission on Federal Paperwork 59
Stephen M. Gershenson

Standards for Electronic Data Interchange 67
Edward A. Guilbert

General Concept of a Data Resource Directory System 71
Clarence E. Hankerson

International Standards Activities of the Economic Commission for Europe 91
Eugene A. Hemley

[1]Papers are sequenced by name of author.

Page

The Federal Information Locator System 97
Forest W. Horton, Jr.

The Urgent Need for Energy Information Standards
(To "Btu" or Not To "Btu") 113
Richard W. Kline, Jr.

Criteria for the Selection of Data Dictionary/
Directory Systems 119
Bernard K. Plagman

Some History of Traditional Trade Documents 135
Robert J. Porter

Data Resource Management 139
Daniel B. Schneider

The Need for Standards in the Health Care Field and
A Sampling of Current Applications 141
Sheila M. Smythe

Appendix A - Participants 143

INDEX OF
KEY WORDS AND PHRASES
TO
AUTHORS' NAMES

ACCREDITED ORGANIZATION METHOD - Duel
AIDS DATA ELEMENT MANAGEMENT SYSTEM - Horton
AIR CARGO - Cavanaugh
AIRLINES - Guilbert
ALIGNMENT - Hemley
AMERICAN ASSOCIATION OF BLOOD BANKS - Brodheim
AMERICAN MEDICAL ASSOCIATION - Duel
AMERICAN NATIONAL RED CROSS - Brodheim
AMERICAN NATIONAL STANDARD FOR BIBLIOGRAPHIC INFORMATION INTERCHANGE ON MAGNETIC TAPE (ANSI Z39.2-1971) - Albert
AMERICAN NATIONAL STANDARDS COMMITTEE METHOD - Duel
AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI) - Cavanaugh; Duel
ANALYTICAL METHODOLOGIES - Kline
ANSI MEDICAL DEVICES STANDARDS MANAGEMENT BOARD - Duel
ARMY RESEARCH AND DEVELOPMENT INFORMATION SYSTEM - Horton
ATTRIBUTES - Hankerson
AUTOMATED AIRPLANE TICKETS - Finnerty
AUTOMATED DATA DICTIONARIES - Albert
AUTOMATIC DATA PROCESSING - Hemley

BANKS - Guilbert
BILL OF LADING - Cavanaugh; Porter
BIOLOGICAL AND PHYSICAL WORLD - Gautier/Shetler/Black
BLOOD BANKING - Brodheim
BLOOD BANKING OPERATIONS - Brodheim
BLOOD PRODUCTS - Brodheim
BUREAU OF MINES (BOM) - Albert

CARGO DATA INTERCHANGE SYSTEM (CARDIS) - Cavanaugh; Hemley; Porter
CATALOGING AND INDEXING SYSTEM - Horton
CATALOG OF FEDERAL DOMESTIC ASSISTANCE - Horton
CERTIFICATE OF ORIGIN - Cavanaugh
CIVIL AERONAUTICS BOARD DATA INDEX - Horton
CLASSIFICATION AND CODING STRUCTURES - Kline
CODABAR OPTICAL BAR CODE - Brodheim
CODING - Hemley
COLLECTIONS - Gautier/Shetler/Black
COMMERCIAL INVOICE - Porter
COMMISSION ON FEDERAL PAPERWORK - Gershenson
COMMISSION'S STUDY OF DUPLICATION - Horton
COMMITTEE FOR COMMONALITY IN BLOOD BANKING AUTOMATION (CCBBA) - Brodheim
COMMON ACCESS REFERENCE NUMBER - Hemley
COMMUNICATION - Finnerty
COMPUTERIZED INTERNATIONAL DOCUMENTATION SYSTEM - Porter
CONTAINERIZATION - Cavanaugh
CONVERSION FACTORS - Kline
CURRENT RESEARCH INFORMATION SYSTEM - Horton

DANGEROUS GOODS - Hemley
DARCOM DATA ELEMENT DICTIONARY - Horton
DATA - Porter
DATA BASE ADMINISTRATION - Plagman
DATA BASE MANAGEMENT SYSTEMS - Plagman

DATA BASES - Gautier/Shetler/Black
DATA CATALOG - Albert
DATA CATALOGUE - Horton
DATA COLLECTION FORM - Albert
DATA COLLECTION/PUBLICATION FORMATS - Kline
DATA DESCRIPTION FILE - Albert
DATA DICTIONARY - Horton
DATA DICTIONARY/DIRECTORY SYSTEM - Horton; Plagman
DATA ELEMENT - Cavanaugh; Gautier/Shetler/Black; Hankerson; Hemley; Kline
DATA ELEMENT DEFINITIONS - Albert
DATA ELEMENT DICTIONARY - Brown; Hankerson; Horton
DATA ELEMENT DICTIONARY/DIRECTORIES - Brown
DATA ELEMENT DIRECTORY - Brown; Schneider
DATA ELEMENT NAMES - Albert
DATA INTERCHANGE - Cavanaugh
DATA ITEMS - Kline
DATA LABEL CONTROL SYSTEM - Horton
DATAMANAGER - Horton
DATA PROBLEMS - Gautier/Shetler/Black
DATA PROCESSING - Gautier/Shetler/Black
DATA PROCESSING MANAGEMENT - Hankerson
DATA QUALITY - Kline
DATA REQUIREMENTS - Hemley
DATA RESOURCE DIRECTORY - Brown; Hankerson
DATA RESOURCE MANAGEMENT - Brown; Hankerson; Schneider
DATA SOURCES - Kline
DATA STANDARDIZATION - Gautier/Shetler/Black
DATA STANDARDS - Finnerty
DATA SYSTEMS - Kline
DATA TRANSMISSION - Finnerty
DB/DC DATA DICTIONARY SYSTEM - Horton
D-20 DATA ELEMENT DICTIONARY - Horton
DEFENSE DOCUMENTATION CENTER SYSTEM - Horton
DEFENSE INTEGRATED DATA SYSTEM - Horton
DEPARTMENT OF ENERGY (DOE) - Albert; Kline
DEPARTMENT OF TRANSPORTATION - Hemley
DICTIONARY OF FEDERAL DATA ELEMENTS - Horton
DIRECTORY OF FEDERAL INFORMATION RESOURCES - Horton
DIRECTORY OF PUBLIC-USE REPORTS AND FORMS - Horton
DISTRIBUTED DATA BASE - Cavanaugh
DOCK RECEIPT - Cavanaugh
DOCUMENTATION STANDARDIZATION - Hemley
DRAFTS - Porter
DRAWBACK FORMS - Cavanaugh
DUPLICATION - Horton

ECONOMIC COMMISSION FOR EUROPE - Cavanaugh; Hemley
ELECTRONIC BANKING - Finnerty
ELECTRONIC DATA INTERCHANGE - Guilbert
ENERGY - Albert; Kline
ENERGY CRISIS - Kline
ENERGY DATA AND COLLECTION - Albert
ENERGY INFORMATION - Albert; Horton; Kline
ENERGY INFORMATION STANDARDS - Kline
ENERGY INFORMATION STANDARDS PROGRAM - Kline
ENERGY INFORMATION SYSTEMS - Kline
ENERGY PROGRAMS - Horton
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION (ERDA) - Albert
ENTITIES - Hankerson
ENVIRONMENTAL INFORMATION SYSTEMS DIRECTORY - Horton
EXPORT DOCUMENTS - Hemley

FEDERAL AGENCIES - Albert
FEDERAL DATA ELEMENT DICTIONARY - Horton
FEDERAL ENERGY ADMINISTRATION (FEA) - Albert
FEDERAL ENERGY INFORMATION LOCATOR SYSTEM (FEILS) - Albert; Horton
FEDERAL GOVERNMENT - Gershenson; Kline
FEDERAL INFORMATION LOCATOR - Horton
FEDERAL INFORMATION PROCESSING SERIES (FIPS) - Horton
FEDERAL INFORMATION REPORTING REQUIREMENTS - Gershenson
FEDERAL INTERAGENCY COUNCIL ON ENERGY INFORMATION (FICEI) - Albert; Kline
FEDERAL POWER COMMISSION (FPC) - Albert
FILE - Hankerson
FLOW OF TRADE - Finnerty
FOREIGN COMMERCE - Finnerty
FOREIGN TRADE - Finnerty
FOREIGN TRADE DATA - Finnerty
FOREIGN TRADE DATA COMMUNICATIONS - Finnerty
FORM - Hankerson
FORMS MANAGEMENT - Hankerson
FORWARDERS - Guilbert
FREEDOM OF INFORMATION ACT - Gershenson; Horton

GENERAL AVERAGE - Porter
GENERAL INFORMATION MANAGEMENT SYSTEM - Horton
GOVERNMENT AND PRIVATE LABORATORIES - Albert
GROCERY STORE SCANNERS - Finnerty

HEALTH CARE - Brodheim
HEALTH CARE CREDIT CARD - Duffy
HEALTH CARE FIELD - Duffy
HEALTH INSURANCE PROGRAM - Duffy
HIGHWAY RESEARCH INFORMATION SYSTEM - Horton
HOTEL RESERVATIONS - Finnerty

IMPORT DOCUMENTS - Hemley
INCREASED PRODUCTIVITY - Cavanaugh
INDEX TO CIVIL SERVICE COMMISSION PUBLICATIONS - Horton
INDUSTRY - Albert
INFORMATION DIRECTORY - Horton
INFORMATION DUPLICATION - Horton
INFORMATION HANDLING - Albert
INFORMATION MANAGEMENT STUDY - Horton
INFORMATION PROCESSING STANDARDS - Horton
INFORMATION REPORTING BURDEN - Gershenson
INFORMATION RQMTS CONTROL AUTOMATED SYSTEM - Horton
INFORMATION TECHNOLOGY - Finnerty
INPUTS - Hankerson
INSURANCE CERTIFICATE - Cavanaugh
INTERMODAL SHIPPING - Cavanaugh
INTERMUSEUM DATA EXCHANGES - Gautier/Shetler/Black
INTERNATIONAL DIRECT DIAL TELEPHONE - Finnerty
INTERNATIONAL MOVEMENT OF FREIGHT - Cavanaugh
INTERNATIONAL STANDARD DOCUMENTATION FORMAT FOR BIBLIOGRAPHIC INFORMATION INTERCHANGE ON MAGNETIC TAPE (ISO 2709-1973) - Albert
INTERNATIONAL STANDARDS ORGANIZATION - Cavanaugh
INTERNATIONAL TRADE DATA STANDARDS - Finnerty
INTERNATIONAL TRADE INFORMATION - Cavanaugh
INVENTORY RECORDS - Porter

LETTERS OF CREDIT - Porter
LEXICON - Horton
LOCATOR SYSTEM - Horton
LOCATOR SYSTEM CONCEPT - Horton

LOGISTICS DATA RESOURCE MANAGEMENT SYSTEM - Horton

MACHINE-READABLE SYMBOLS - Brodheim
MANAGEMENT - Duffy
MANUFACTURERS - Guilbert
MEDICAL LITERATURE ANALYSIS AND RETRIEVAL SYSTEMS - Horton
MOTOR CARRIERS - Guilbert
MUSEUMS - Gautier/Shetler/Black
MUSEUMS DATA SYSTEMS - Gautier/Shetler/Black
MUSEUM RECORD SYSTEMS - Gautier/Shetler/Black
MUSEUM SPECIMENS - Gautier/Shetler/Black

NATIONAL BUREAU OF STANDARDS - Cavanaugh
NATIONAL COMMITTEE ON INTERNATIONAL TRADE DOCUMENTATION (NCITD) - Cavanaugh; Hemley
NATIONAL HEALTH INSURANCE LEGISLATION - Duffy
NATIONAL HEALTH PLANNING GUIDELINES - Smythe
NATIONAL HISTORY INFORMATION SYSTEM - Horton
NATIONAL INFORMATION REFERRAL SYSTEM - Horton
NATIONAL REFERRAL CENTER DIRECTORY - Horton
NATIONAL SCIENCE FOUNDATION (NSF) - Albert
NATIONAL SCIENCE INFORMATION EXCHANGE - Horton
NATIONAL TECHNICAL INFORMATION SERVICE SYSTEM - Horton
NATURAL HISTORY - Gautier/Shetler/Black
NATURAL HISTORY MUSEUMS - Gautier/Shetler/Black
N.B.S. SPECIAL PUBLICATION - Brown
NETWORKS - Gautier/Shetler/Black
NOMENCLATURE - Kline
NUCLEAR DESALINATION INFORMATION CENTER SYSTEM - Horton

OCEAN CARRIERS - Guilbert
OFFICE OF FACILITATION - Hemley; Porter
OFFICIAL DIRECTORY OF MUSEUMS - Gautier/Shelter/Black
OPTICAL BAR CODES - Brodheim

PAPERWORK - Horton
PATIENT INFORMATION - Duffy
PRIVACY ACT - Gershenson; Horton
PROCESSES - Hankerson
PRODUCT IDENTIFICATION CODE - Brodheim
PROGRAM MANAGEMENT - Hankerson
PUBLIC - Albert
PUBLIC LAW 93-556 - Gershenson
PUBLIC USE REPORT PROFILE - Horton

RAILROADS - Guilbert
RECORD - Hankerson
RECORDS ASSOCIATION SYSTEM II - Horton
RECORDS MANAGEMENT - Hankerson
REPORT - Hankerson
REQUIREMENTS DEFINITION - Brown

SCIENTIFIC AND TECHNICAL INFORMATION SYSTEM - Horton
SHIPPERS - Guilbert
SHIPPER'S EXPORT DECLARATION - Cavanaugh; Porter
SOFTWARE SELECTION - Brown
SPECIAL INTEREST DATA - Albert
STANDARD DATA ELEMENTS AND DEFINITIONS - Albert
STANDARDIZATION - Albert; Brodheim
STANDARDS - Duel; Duffy; Finnerty; Guilbert; Smythe
STANDARDS IN THE HEALTH CARE FIELD - Smythe
STANDARDS' PROCEDURES AND MANAGEMENT SERVICES - Duel

STANDARDS PROGRAMS - Kline
STATE AND LOCAL GOVERNMENTS - Albert
SUBJECT UNDER STUDY - Hemley

TECHNICAL INFORMATION EXTENSION SYSTEM - Horton
TECHNOLOGICAL INNOVATION - Cavanaugh
TECHNOLOGY TRANSFER - Finnerty
TERMINAL - Duffy
TERMINOLOGY - Kline
TRADE DATA - Porter
TRADE DATA STANDARDS - Finnerty
TRADE DOCUMENTS - Porter
TRANSPORTATION - Cavanaugh; Finnerty
TRANSPORTATION DATA COORDINATING COMMITTEE (TDCC) Cavanaugh; Guilbert
TRANSPORTATION NOISE RESEARCH INFORMATION SERVICE - Horton

UCC TEN DATA DICTIONARY - Horton
UNITED NATIONS - Cavanaugh
UNITED STATES GEOLOGICAL SURVEY (USGS)- Albert
UNITED STATES MERCHANDISE TRADE BALANCE - Finnerty
UNITED STATES STANDARD MASTER - Cavanaugh
UNITS OF MEASURE - Kline
UNIVERSITIES - Albert

VETERANS ADMINISTRATION DATA ELEMENT DICTIONARY - Horton
VOLUNTARY CONSENSUS STANDARDS - Duel

WATER DATA SYSTEM - Horton
WATER RESOURCES SCIENTIFIC INFORMATION CENTER SYSTEM - Horton
WORKING PARTY ON FACILITATION OF INTERNATIONAL TRADE PROCEDURES - Hemley

FOREWORD

The Third National Symposium on the Management of Data Elements in Information Processing was held at the National Bureau of Standards on September 28 through 30, 1977. This Symposium was jointly sponsored by NBS and American National Standards Institute Committee X3L8 on the Representation of Data Elements. Several hundred professionals attended sessions devoted to the recognition of data as a resource that needs to be harnessed, costed, and managed.

Primary emphasis was given to the subject of Data Resource Management, and was based on the assumption that it is data that is being collected, processed, published, filed, transmitted, used and misused. It is data, man's oldest resource that is causing paperwork management problems. It is data that is demanding more and more of the organization's resource dollar for capital investment. It is data, the raw material, that continues unmanaged, uncontrolled, and the subject of today's privacy issues.

This Symposium featured the highlights of the Findings and Final Report of the Commission on Federal Paperwork to the U.S. Congress, the Federal Information Locator System, data resource management and data resource directories. Also considered were standards needs in the areas of energy information systems, the health care field, national and international trade data standards, and standards for museum data interchange systems.

NBS is pleased to have the opportunity of making the text of the papers, presented at the Symposium, available in these proceedings. We feel that these are representatives of the state-of-the-art of current data management practices and should provide a useful base for further activities in this important field.

However, it must be stressed that the responsibility for the content of the papers provided in these proceedings rests with the individual authors and their organizations and does not reflect endorsement by the National Bureau of Standards.

Hazel E. McEwen, Program Chairperson
Office of ADP Standards Management
Institute for Computer Sciences and
Technology, NBS

FIRST SESSION: INFORMATION MANAGEMENT AND THE FEDERAL INFORMATION LOCATOR SYSTEM

Program Co-Chairperson: Walter Schlenker
General Electric Company

Session Chairperson: William H. Kenworthey, Jr.
Department of Defense

Participants:

Stephen M. Gershenson
Commission on Federal Paperwork

Forest W. Horton, Jr.
Commission on Federal Paperwork

From left to right: Walter Schlenker, William H. Kenworthey, Jr., Stephen M. Gershenson, Forest W. Horton, Jr.

Note: Addresses of speakers are in Appendix A.

SECOND SESSION: STANDARDS FOR APPLICATION SYSTEMS (ENERGY INFORMATION)

Program Co-Chairperson: Walter Schlenker
General Electric Company

Session Chairperson: Nevaire Serrajian
Department of Energy

Participants:

Ted M. Albert
Department of Energy

Richard Kline
Federal Energy Administration

From left to right: Walter Schlenker, Nevaire Serrajian, Ted M. Albert, Richard Kline

SECOND SESSION: STANDARDS FOR APPLICATION SYSTEMS (HEALTH CARE)

Panel Chairperson: Sheila M. Smythe
Blue Cross-Blue Shield of
Greater New York

Participants:

Eric Brodheim, Sc.D.
American Blood Commission

Ward Duel, MPh
American Medical Association

Gerald J. Duffy
Blue Cross Association of
Chicago

From left to right: Sheila M. Smythe, Eric Brodheim, Gerald J. Duffy, Ward Duel

SECOND SESSION: STANDARDS FOR MUSEUM DATA INTERCHANGE SYSTEMS

Panel Chairperson: Charles Kenny
U.S. Civil Service Commission

Participants:

Craig Black
Carnegie Museum of Natural History

Gary Gautier
Smithsonian Institution

Stanwyn G. Shetler
Smithsonian Institution

From left to right: Gary Gautier, Stanwyn G. Shetler, Craig Black

THIRD SESSION: INTERNATIONAL TRADE DATA STANDARDS

Chairperson: Walter Schlenker
General Electric Company

Participants:

Peter J. Finnerty, Esq.
Sea-Land Services, Inc.

Robert J. Cavanaugh
General Electric Company

Eugene Hemley
National Committee on International Trade Documentation

Edward A. Guilbert
Transportation Data Coordinating Committee

Robert J. Porter
Eastman Kodak Company

FOURTH SESSION: DATA DICTIONARY/DIRECTORY SYSTEMS

Bernard K. Plagman
DBD Systems, Inc.

DATA RESOURCE MANAGEMENT

Panel Chairperson: Aaron Hochman
Office of Secretary of Defense

Participants:

Paul Brown
U.S. Civil Service Commission

Daniel Schneider
U.S. Department of Justice

Clarence Hankerson
Federal Aviation Administration

From left to right: Walter Schlenker, Aaron Hochman, Paul Brown, Daniel Schneider, Clarence Hankerson

FIFTH SESSION: WORKSHOPS

Harry S. White, Jr. (Coordinator) - Workshop on Establishing a Data Standardization Effort.

George Begnal, General Electric Co. - Demonstrating a Trade Data Interchange Application.

TG-17 Group - Data Resource Management/Data Resource Directories
(Standing) left to right: Daniel Schneider, Aaron Hochman (Chairperson), Thomas Chittenden, William Morgan, (Seated) left to right: Roland Page, Robert Mattes, Clarence Hankerson, Paul Brown

Energy Information - Top Priority For Today

T. M. Albert
Director
Office of Environmental Information Systems
U. S. Energy Research and Development Administration
Washington, D. C. 20545

The OPEC oil embargo of 1973, in part, created widespread concern over the total energy picture in the United States. The embargo created the impetus toward developing short and long-term plans and policies regarding every aspect of the Nation's energy problems. To attempt to define the extent of the energy problem, one of the first tasks undertaken was to gather as much information as possible.

At that time there was no national energy information program. Various Federal offices collected data to suit their special needs and developed unique methods and systems. Little attempt was made at coordination with other similar activities.

After the embargo, Federal energy-related activities in collection, analysis, and dissemination of data continued at a greater rate. Speed and expediency took precedence over control and coordination. Efforts were fragmented, the language of energy data and collection was imprecise and uncontrolled, and there were wide variances in reported data values as well as in methods of dissemination.

Today our energy situation has not improved appreciably. We now import more than half the oil we use which brings with it the expected damaging economic effects. The diagnosis of the U.S. energy problem can be stated simply: demand for energy is increasing while the available domestic supplies of oil and natural gas have been declining. The cure is far from simple, and its solution requires attention to complex economic, political, and social factors, as well as technical and scientific matters.

Similarly, we need and continue to collect greater quantities of data and information which cover a wider and wider scope. However, at least until relatively recently, there has been little movement toward methods of control, coordination, standards, and validation of the data. The burden on respondents to requests for data is crushing. At the same time, because of the energy problem, information is one of our top priorities.

To better understand our need for and the difficulties involved in applying controls and standards, and better management of information handling, lets look at the scope of the problem. One of the biggest information-intensive areas today is that revolving around energy, with the largest part of the effort being carried on in the Federal Government. More than 40 different bureaus or agencies collect data through more than 200 separate programs. And this does not include R&D type information.

Energy information collected and processed in the Government serves three basic needs: policy, regulation, and R&D. The bulk of the collected data is economic or statistical which creates many numeric data bases. The rest is, of course, bibliographic or text material.

The data and information collected concerns every related aspect and all possible sources of energy--supply and demand, economics, pricing, companies' operations and finances, resources and reserves, production, transportation, imports, transmission, socioeconomics, ecology, environment, health, safety, conservation, specific technologies such as solar, nuclear, geothermal, and so on. Sources of the data are: Federal agencies, state and local governments, industry, government and private laboratories, universities, the public, and special interest groups, With the exception of the public, all of these groups also collect data. Literally hundreds of data collection forms are involved; there is much redundancy, duplication, and a lack of standard definitions.

If we will accept the fact of the existence of the energy problem with all its ramifications, we should not have to justify the value or need of the information; it should be clear. The following list of necessary information application areas should be enough to illustrate our strong need for a wide range of data: crises--brownouts, coal strikes, embargo; nuclear--fuel management/use, safety, waste management, reactor construction; commercialization of technologies; solar heating and cooling use and demand; policy, planning, regulations, decisionmaking; environmental tradeoffs; reserves estimates; electricity demand; etc.

To those of us in the data and information business the need for a range of standards in this kind of environment is clear. As an example, consider the terms "resources" and "reserves" and their definitions. Actually, in the world of energy, these terms are not clearly defined. If you cannot clearly measure a reserve of a resource how do you define it? When does a reserve become a resource or vice versa? Are they the same thing? Is a fuel reserve everything, or everything that can be recovered, or everything that is economically recoverable? How do you measure reserves--a different measure for each type or some common equivalency measure?

Consider the phrase "crude oil production". Do we measure it when it comes from the well or when it transfers to a storage tank or pipeline or refiner? Do we include in the measure the sediment that may be present or water that may be mixed in?

Take the term "imports". For purposes of statistical analysis, forecasting, and reporting, when does crude oil become an import? When it is loaded in a contracted tanker? When the tanker arrives in U.S. waters or a U.S. port? Or when it is unloaded?

Some of these questions may seem trivial but they are extremely important. Unfortunately there are thousands of terms that are defined differently by different users and data collectors. Some have different names but mean the same thing. The lack of standardization causes a great burden and workload for the respondents of questionnaires, particularly from the Federal sector. The

problems of reporting to the President, the Congress, and the people are magnified. We have all seen examples of different quantities of a particular fuel production or a reserve being reported by different Federal groups.

The word standards alone covers a multitude of interests as regards energy: safety, regulatory, data elements, computer/communications associated work, scientific and technical terminology, building materials, weights and measures, environmental, etc. As you all know, the problems of standardization are further compounded because of the many diverse groups who must agree. Also, there are two kinds of data and information with which we are concerned--text and numeric.

How are we handling energy data and information today? What steps are we taking and what should we be doing towards the end of standardization which in turn will make easier our overall task of solving the Nation's energy problems?

The bulk of energy-related numeric data is collected by the Federal Energy Administration (FEA), the Bureau of Mines (BOM) and the Federal Power Commission (FPC). Most of it is collected through the use of data collection forms which may number close to 1000. Submission of many of these forms is required by law, the others are submitted voluntarily. The number of data elements involved runs in the thousands.

The Energy Research and Development Administration (ERDA) contributes the bulk of R&D or text-type information with additional amounts coming from the National Science Foundation (NSF), the U.S. Geological Survey (USGS), and some others; this information derives from citations, abstracts, and bibliographies. ERDA alone adds over 150,000 citations and abstracts a year just to its main energy data base (there are over a million citations in the file now).

Most of the data, numeric or text, is input to computer readable files. Unfortunately, in most cases, many of the files are unique, that is, there is a lack of standards and standard methods. Any comparison, validation, and cross-checking is extremely difficult if not impossible. Items have the same name but different definitions and vice versa.

However, we are beginning to see an understanding of these problems and attempts to control the proliferation of data bases through the use of data element dictionaries and data base management systems. Both FEA and FPC have automated data dictionaries which contain descriptions of their data collection forms, their systems, as well as the names of the data elements and their definitions. These dictionaries are of two types: (1) a stand-alone dictionary, and (2) a dictionary that helps drive the system.

ERDA has developed a subject thesaurus of scientific and technical terms and is presently updating it. A data catalog and index to all energy-related data bases and models available in ERDA is being prepared. ERDA has funded an Interlaboratory Working Group on Data Exchange (IWGDE) (Brookhaven, Los Alamos, Oak Ridge, etc.) which, as one of its accomplishments, is establishing standards for exchange of machine readable data between the laboratories. The definition phase has been completed and implementation is ongoing of a hardware and content - independent standard and process for exchange of data via magnetic tape. A single

software program at each lab is being developed to create and interpret the tape structure and labels; transfer and conversion of data will in the future be much easier. The IWGDE is also involving itself in standards for geocoding and consideration of documentation of existing ERDA graphics techniques.

For those interested in more detail, the proposed data exchange standard was developed by the IWGDE as a generalization of the American National Standard for Bibliographic Information Interchange on Magnetic Tape (ANSI Z39.2-1971) and the International Standard Documentation - Format for Bibliographic Information Interchange on Magnetic Tape (ISO 2709-1973). The basic structures defined by these standards were extended to accommodate the interchange of a wide variety of information. In order to make the interchange files content - independent, a Data Description File (DDF) was added. The file contains control parameters and data descriptions necessary to interpret data records and data elements. Thus a magnetic tape embodying this structure consists of three parts: (1) The Data Description File (DDF), (2) The Data File (DF), and (3) The required tape labels.

FEA has developed an inventory called the "Federal Energy Information Locator System". A document under that title was published which lists the more than 40 bureaus and agencies, mentioned earlier, that have the over 200 different energy-related programs.

Some of you may be familiar with the Federal Interagency Council on Energy Information of which I am presently chairman. The Council consists of representatives of 15 Federal agencies. We are concerned with policy and problems that concern all aspects of energy information--how can we reduce duplication, how can we share, how can we provide the right data on which to make decisions? Several working groups are sponsored--one is concerned with the question of what and how much energy information is really necessary in order for the Government to make policy and decisions; another carried through a successful pilot program on the feasibility of taking an inventory of all energy statistical and economic data collected in the Federal Government (the project has now been referred to FEA for full implementation); another has looked at the problem of proprietary data; and another has been establishing the foundation of an extensive energy data standards program. The standards project has just been referred to the new Department of Energy (DOE) for implementation.

Let me say a word here regarding the pilot program concerned with the feasibility of taking an inventory of all energy statistical and economic data collected by the Federal Government. At first glance this might seem an impossible task considering the many agencies involved with their proliferation of programs.

The concept for the overall task involved a three-phase approach: (1) collect all data collection forms; (2) enter into an automated file all the data element names with detail about the forms; and (3) collect and enter existing data element definitions. The file could then be analyzed for duplication and redundancy, and provide us a foundation for establishing standard data elements and definitions.

The pilot program utilized approximately 90 forms from four different agencies. It was quite successful and indicated a "doable" task in a reasonable length of time with a reasonable amount of resources. As I mentioned, the project was referred to FEA, which began the work. This work will continue under the Energy Information Administration group in the Department of Energy.

Let me talk a bit about the new Department. DOE officially comes into being on October 1. Many of the agencies I mentioned will become part of DOE--ERDA, FEA, FPC, and parts of BOM, DOI and DOD. By bringing them together we will have grouped under one roof the main Federal elements that are responsible for the Nation's energy policy, and implementing efforts to solve our energy problem. I think this will make the job easier.

There will be two main information programs in DOE--the Energy Information Administration (EIA), and the Office of Technical Information (OTI). OTI will concern itself with scientific, technical, and engineering information. EIA will concern itself with statistical and economic supply and demand related information. There will also be various specialized information programs associated with the DOE technological programs.

The standards project referred from the Interagency Council will be undertaken by the EIA. It will involve standard terms, definitions, weights and measures, and the promulgation of these standards. The OTI will work closely with EIA concerning itself with work on standard terms, definitions, and measures related to scientific and engineering data because of its experience and on-going work in this area. The Interagency Council will coordinate and press for the adoption of developed standards and help promulgate their use. Also, the work we are discussing will be coordinated with the National Bureau of Standards.

What I have described is a possible framework and plan on which to base the energy information standardization program. A great deal of effort has gone into developing this foundation and in bringing it to the attention of top officials. Work has already begun which you will soon be hearing more about.

Along with this, in DOE, will be an expansion of various efforts to develop knowledge of existing data and its quality; and to better manage its collection, processing, and dissemination. DOE's data and information needs will range from details of petroleum and natural gas supply and demand, to solar commercialization, to environmental impacts in Alaska, to waste management, to water pollution, to transportation, and so on. Data quality and dissemination will be stressed to allow much wider use by diverse audiences. As we all know, the job will be a difficult but necessary contribution to a solution to the problem of an adequate, continuing, and economically feasible energy supply for this country.

Standardization of Data Elements and Machine-Readable Symbol to Facilitate Blood Banking

Eric Brodheim, Sc.D.

Chairman, Committee for Commonality in
Blood Banking Automation
American Blood Commission

Standardization of data elements describing blood products was an important requirement in introducing machine-readable technology in order to improve the safety and efficiency of blood banking operations. This paper discusses some features of this effort and some of the implications for the broader hospital patient care system.

Key words: Blood Banking; machine-readable symbols; optical bar codes; standardization.

1. Introduction

As automated equipment is being introduced to improve the safety and the economics of health care, there is a growing need to standardize data elements and the symbols for representing these in machine-readable form. The need in blood banking for this was recognized about three years ago by the blood service community consisting of approximately six thousand establishments that collect, process and/or transfuse blood in the United States. This paper discusses some features of this effort and some of the implications for the broader hospital patient care system of which blood banking is only one small sector.

The three national organizations that broadly represent the blood service community; namely, the American Association of Blood Banks, the American National Red Cross and the Council of Community Blood Centers – jointly organized the Committee for Commonality in Blood Banking Automation (CCBBA) to undertake such a standardization effort. This project, which has continued under the aegis of the recently created American Blood Commission, has developed and tested an approach for standardizing data elements pertaining to blood products and their associated records in a compatible eye-readable and machine-readable form.

2. Background Information

As each of the over 10 million annual whole blood donations is made, a unique identification number is assigned to the "unit" drawn, and is affixed to each of the associated blood bag labels, each of the pilot tubes containing blood samples for testing purposes and for each of the donation records.

A blood unit may be transfused as whole blood or can be divided into components. The primary therapeutic components of blood are the red blood cells, which are clinically useful for 21 days after collection; the plasma which, if frozen following collection is usable for two years; platelets with a useful life of 48 to 72 hours; leukocytes which have a clinically useful life of approximately 6 hours and cryoprecipitate, containing the clotting factors, which can be separated from the plasma and kept frozen until used.

During the time scale (usually measured in hours) required to separate the various components and to distribute them to patients, a number of tests must be performed on the blood samples that were collected at the time of donation to check for indications of hepatitis, syphilis or unusual antibodies. The donor records may also be processed against registries of donors whose blood is, for various reasons, unsuitable for transfusion purposes. While all of this is done the individual blood components must be labeled to indicate their unique identification number, their ABO and Rh, the results of these various tests and checks, the type of product and its expiration date.

Whole blood and red blood cells, which account for the vast majority of transfusions are then delivered to transfusion services where they are maintained for compatibility checking (cross-matching) for possible transfusion to individual recipients. On the average, three such compatibility tests are performed and recorded before the unit is transfused.

In recording compatibility tests, both the unique identification of the blood donation (linking back to the identification of the donor of the originating whole blood unit) as well as that of the possible recipient needs to be preserved and associated. In transfusion services where there is a significant volume of compatibility checking there is always a possibility of a mix-up whereby the wrong blood unit or a unit that has not been fully tested for compatibility is delivered for transfusion to a patient. Despite elaborate manual checks to prevent such occurrences, there are human lapses which cause extremely serious and even fatal results.

The desire to improve safety and to reduce the labor and cost of the necessary record keeping has led the blood banking community to seek the use of automation. This has brought about the need to standarize the data elements involved. While most of this effort has currently been in the processing of blood it will invariably move through the transfusion service to the patient's bedside where it will be used to verify that the proper blood products are being administered to the patient. This has obvious parallels to the identification and the clinical testing of all types of samples from patients leading to the verification of proper medications being administered.

3. Standardization Effort in Blood Banking

The data elements pertaining to a blood product that have been standardized are the following:

(1) the unique identification of the donation;

(2) the identification of the specific product including volume and expiration date;

(3) the identification of the results of the testing and screening performed on the blood sample and on the donor.

The manner in which these data elements have been standardized and are represented on blood products in machine-readable form is discussed below.

The physical placement of the labels containing these data elements is performed in stages, and is illustrated in figure 1. The unique unit number

identification is applied at the time that the whole blood donation is made. The label identifying the product, separated into a particular satellite bag, is applied once that product has been separated. The testing - disposition label is affixed after all laboratory testing and registry checking is completed and/or the unit is released for possible transfusion.

The placement of the above data elements in machine-readable form on a fully labeled blood product is illustrated in figure 2. The specific placement of machine-readable identification on the bag was selected to facilitate compatibility between manual and automated operations and to accommodate the many diverse manual and automated protocols in use.

3.1. Unique Identification of Donation

In order to uniquely identify a unit of blood within the United States two consecutive pieces of information are necessary. The first identifies the location where the donation was made. The second is a donation number which is unique within that location.

Each U.S. blood drawing location is identified by a unique five-digit number which is assigned by the FDA's Bureau of Biologics. This is preceded by a two-digit code which identifies the drawing location as being within the United States making up a seven-digit registration number. Informal discussions are underway for other countries to coordinate the assignment of the two-digit "region" codes internationally.

In all cases the machine-readable donation number will be a seven-digit number which coupled with the seven-digit registration number gives a fourteen-digit number that will be unique throughout any part of the world adopting this system. For the benefit of the many smaller collection centers there is the option to eliminate leading zeros in eye-readable form. Since many blood centers utilize alphanumeric designations (which are felt to be desirable for manual use) provision is made for the first two digits to be represented by one or two alphabetic characters in accordance with Table 1.

The American Red Cross has 57 regional blood centers, collects approximately 50% of the blood in the U.S. and operates under a single national license. All of these centers have individual registration numbers which are differentiated in machine-readable form. However, in order to maintain compatibility with existing manual systems and to fit in with the identification scheme for the American Red Cross, a two-digit center code (which redundantly conveys the same information as the unique registration number for that center) is added at the beginning of the unit identification number in eye-readable form. This is illustrated in figure 3 which shows the various options for unit number representation.

Figure 3 also gives examples of some of the different unit number configurations and the way these are organized in pads. Each of the 50 sheets of a pad corresponds to one unique set of identification numbers. These are either in eye- and machine-readable form suitable for use on sample test tubes, on blood bags or on machine-readable donor records; or they are only in eye-readable form suitable for use on donation records and on test tubes where machine readability is not required.

3.2. Identification of Product

The product identification code is made up of five characters. The first three characters are used to identify the particular product. The fourth character identifies the anticoagulant in which the blood is collected or the method of preparation of a component. The fifth character on an empty bag identifies the type of container in which whole blood is collected

Table 1. Alphabetic unit number prefix option

		FIRST LETTER - Primarily used to designate centers in a region.				
		Blank	F	G	K	L
SECOND LETTER - Primarily used to designate mobiles from a given center.	C	00	20	40	60	80
	E	01	21	41	61	81
	F	02	22	42	62	82
	G	03	23	43	63	83
	H	04	24	44	64	84
	J	05	25	45	65	85
	K	06	26	46	66	86
	L	07	27	47	67	87
	M	08	28	48	68	88
	N	09	29	49	69	89
	P	10	30	50	70	90
	Q	11	31	51	71	91
	R	12	32	52	72	92
	S	13	33	53	73	93
	T	14	34	54	74	94
	V	15	35	55	75	95
	W	16	36	56	76	96
	X	17	37	57	77	97
	Y	18	38	58	78	98
	Z	19	39	59	79	99

Example: 04 translates to H
0413897 translates to H13897
6413897 translates to KH13897

(i.e., in a single bag to be used for whole blood, a double bag to be separated into whole blood and plasma, or bags with more compartments for sterile separation into up to five separate transfusable components). When the bag contains 2 separated components, the fifth character indicates the volume of the contained product. This product code logic is summarized in Table 2.

The date of collection is also required since, together with the identification of the product, it defines the expiration date. Typical product labels are shown in figure 4.

Table 2. Product code assignment logic

	DIGITS 1-3	DIGIT 4	DIGIT 5
Whole Blood & Red Blood Cell Products Pre-printed on Primary Bag Label	Product Code	Anticoagulant	Package Code
Empty Satellite Bag Pre-printed on Satellite Bag Label	000	0	Package Code
Whole Blood & Red Blood Cell Product Labels	Product Code	Anticoagulant	Volume Code
Component Product Labels	Product Code	Method of Preparation Code	Volume Code
Blank Product Label	099	0	1

3.3. Identification of Testing and Screening Results

In the vast majority of cases the testing - disposition information is conveyed by a label that indicates that the unit meets all FDA testing requirements and gives the blood group and type (see figure 5). There are a number of exceptions. There is provision for the "emergency release" of products without all tests or checks being completed, in which case the tests done are indicated manually on an emergency release label. In situations where blood is donated by a patient in advance of anticipated surgery and is to be reinfused into the same individual there is provision to so identify the unit by another special label. There are also labels for identifying units that are defective in some manner and therefore not suitable for normal transfusion; or units where tests results indicate the possiblity of hepatitis, syphilis or antibodies and are therefore being held for further testing. Finally, there are special labels for use on units detected to be positive for hepatitis or syphilis which must be suitably disposed of.

4. Machine-Readable Features

In automated operations it is necessary to provide security that the proper information is being read. For this reason, unique control codes are placed before and after all data elements. The control code assignments are shown in Table 3.

In some instances it is necessary in the interest of safety to require that certain data elements be read with a single pass of the light pen. For instance, it is important that unit number and blood type are read together to provide ultimate assurance that the bag is properly labeled. This is of utmost importance since if it is mislabeled this can cause a fatal transfusion reaction. Other data is placed to be read at the same time in the interest of efficiency and convenience.

Table 3. Control code assignments for the CCBBA simplified blood bag labels.

START CONTROL CODE	DATA	STOP CONTROL CODE	USED FOR	USED ON
a0	7 digit FDA Reg. #	1b	Numeric Unit Number	Bag Label
a1	7 digit FDA Reg. #	1b	Alphanumeric Unit Number	
a$	3 digit dating method*	d	Machine-readable Collection Date	Collection Time Label
a8	5 digit dating method**	d	Machine-readable Collection Time and Date	
a0	5 digit blood product code number	2b	Blood Product on Bag Label	Bag Label
a0	5 digit blood product code number	3b	Blood Product on Product Label	Product Label
d	xx***	$b	No Typing Performed	Bag Label
d	51	0b	O Positive	Grouping Labels
d	62	0b	A Positive	
d	73	0b	B Positive	
d	84	0b	AB Positive	
d	95	0b	O Negative	
d	06	0b	A Negative	
d	17	0b	B Negative	
d	28	0b	AB Negative	
d	39	0b	Hold for Further Processing	Grouping Labels
d	40	0b	For Autologous Use Only	
d	4-	0b	Not For Transfusion	
d	5$	0b	Biohazard	
d	7+	0b	Emergency Use Only	
d	7 digit unit number	d	Identification of Blood Bags and Sample Tubes	Unit Number Label

* Julian 3 digit method for indicating day of year

** The first 3 digits use Julian method for indicating day of year and the last 2 digits indicate hour of day in military time.

*** xx = 2 digits: the first indicates anticoagulant code and the second digit indicates the package code.

The CODABAR optical bar code was selected primarily because it seemed to offer the maximum security and had the capacity for being added to in stages. A modification was made to add a "pause" code to indicate the beginning and end of a specific label. Early tests show that without this unreliable reading was obtained especially as labels became worn. With this logic, a bag is always machine-readable regardless of its state in labeling.

5. Remaining Problems

The emphasis to date has been in the automation of the blood collection, processing and distribution operations. The next phase is to standardize those data elements pertaining to the operation of a transfusion service. This will lead to the last phase of making provision for the automation of the bedside transfusion operations. This gives the ultimate safety payoff of verifying that the proper blood products are being transfused.

These steps must be performed in sequence. It is not feasible for the transfusion service to automate unless they receive blood labeled in a common machine-readable form from their one or more supplying blood centers. It is further not feasible to automate at the bedside unless the transfusion service is automated so that it is able to record the blood product-patient identifications and compatibility testing results and send blood units to the bedside with a machine-readable identification that can be matched against the identification of the patient.

It is at this point that the transfusion service becomes an integral part of the patient care system of the hospital. It is clearly undesirable to identify a hospital patient in one way for blood banking operations and in another way for other aspects of the patient care systems. With unit dose packaging of medication being widely implemented it is clear that the same concept of checking to insure that the proper blood products are being administered can be applied to insure the proper administration of any medication. Similarly, the need to insure the integrity of the blood samples taken from a patient for testing blood compatibility is similar for samples taken from the patient for any other type of testing.

CCBBA has attempted to enlist the support of the National Bureau of Standards (NBS), the American Hospital Association (AHA) and other interested parties in controlling the indiscriminate introduction of a multitude of machine-readable codes into the hospital environment. NBS did co-sponsor a very useful meeting at this facility but never followed up on this. Liaison representatives from the American Medical Association and the AHA have taken part in CCBBA meetings. While some constructive interaction has resulted, the AHA has so far not responded to suggestions by the American Blood Commission to engage in a broad based effort of setting criteria for data elements and for machine-readable symbols in the overall patient care system.

Faced with a clear and present need, the CCBBA could not wait an indeterminate period for the evolution of broad-based patient related identification criteria but had to carry out its commitment to introduce a system suitable for blood banking. Before too many years pass, its efforts will be evident at the bedside of the patient. There is legitimate concern for compatibility with other potential patient identification systems. It is in the public interest that this problem should be addressed and we feel that it would be appropriate that the National Bureau of Standards exercise a position of leadership in bringing this about.

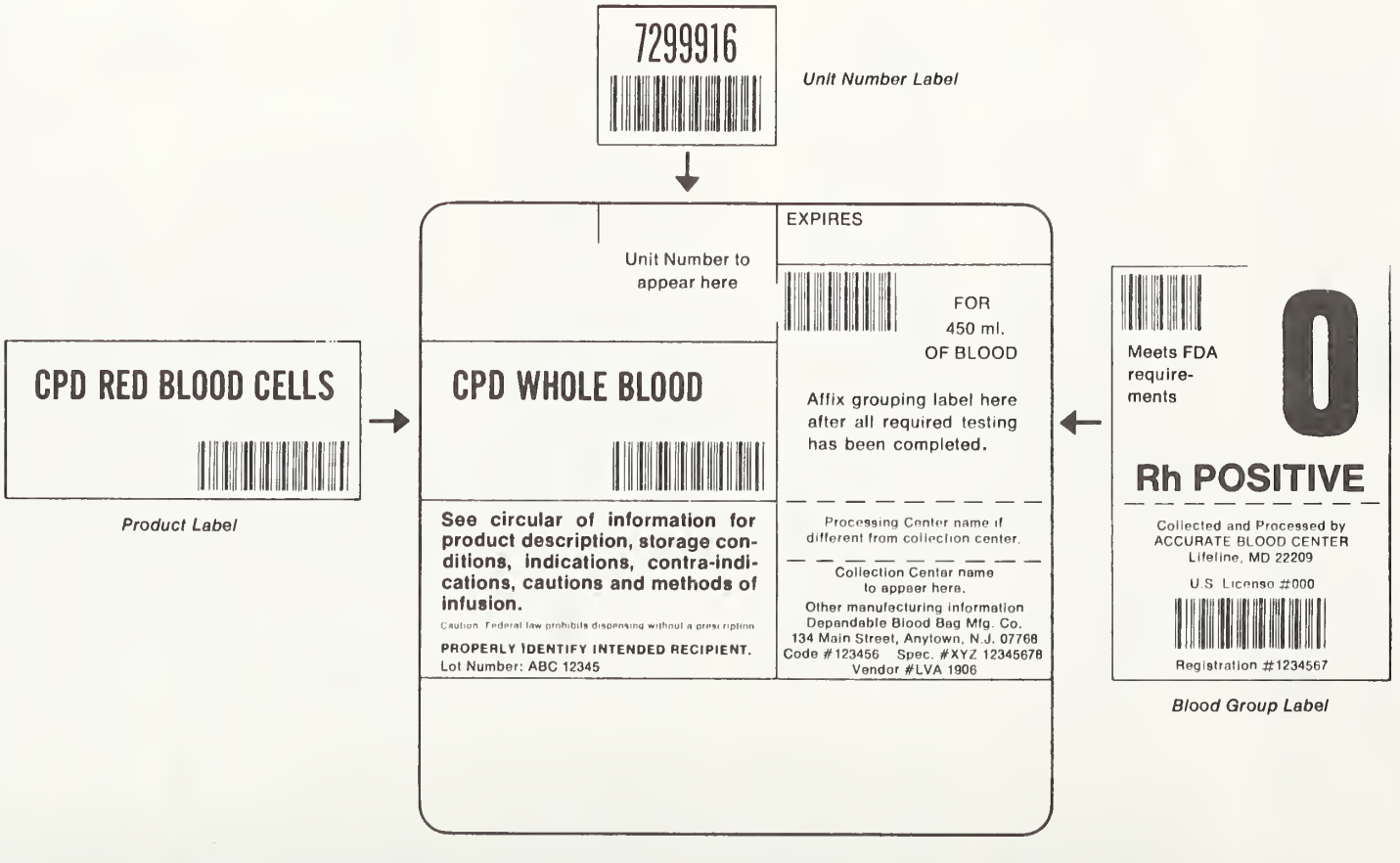

7299916
Unit Number Label
Unit Number to appear here
EXPIRES
FOR 450 ml. OF BLOOD
Affix grouping label here after all required testing has been completed.
CPD RED BLOOD CELLS
Product Label
CPD WHOLE BLOOD
See circular of information for product description, storage conditions, indications, contra-indications, cautions and methods of infusion.
Caution: Federal law prohibits dispensing without a prescription
PROPERLY IDENTIFY INTENDED RECIPIENT.
Lot Number: ABC 12345
Processing Center name if different from collection center.
Collection Center name to appear here.
Other manufacturing information
Dependable Blood Bag Mfg. Co.
134 Main Street, Anytown, N.J. 07768
Code #123456 Spec. #XYZ 12345678
Vendor #LVA 1906
Meets FDA requirements
O
Rh POSITIVE
Collected and Processed by
ACCURATE BLOOD CENTER
Lifeline, MD 22209
U.S. License #000
Registration #1234567
Blood Group Label

Figure 1. Placement of labels on satellite bag label

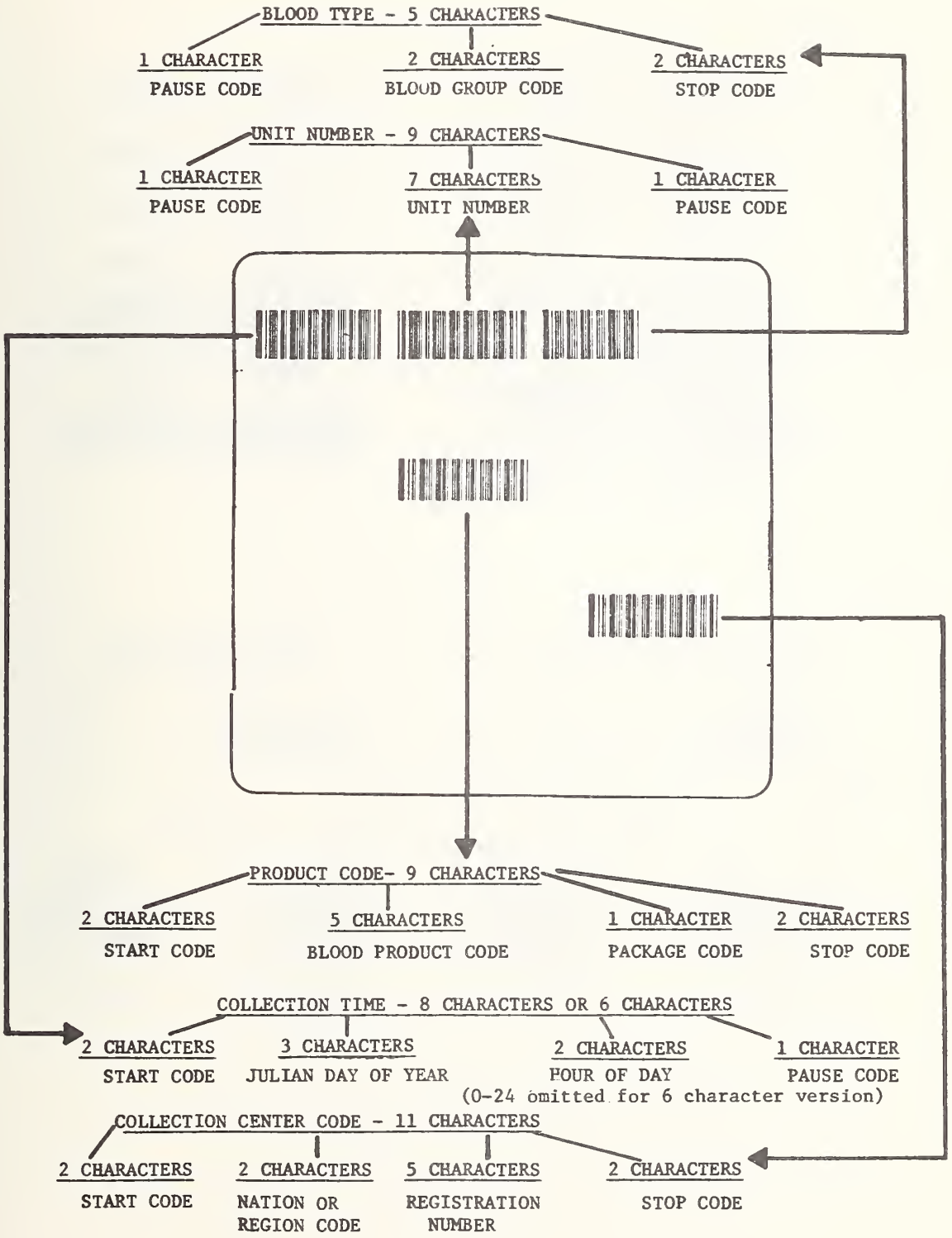


Figure 2. Placement of Machine-Readable Information on Blood Bags

Numbering Variations on Unit Number Labels

Zeros suppressed for smaller centers

Red Cross Centers

Alpha characters to conform with existing systems

Full Numbers for larger centers

Figure 3. Unit Number Set Variations

Figure 4. Representative Product Labels

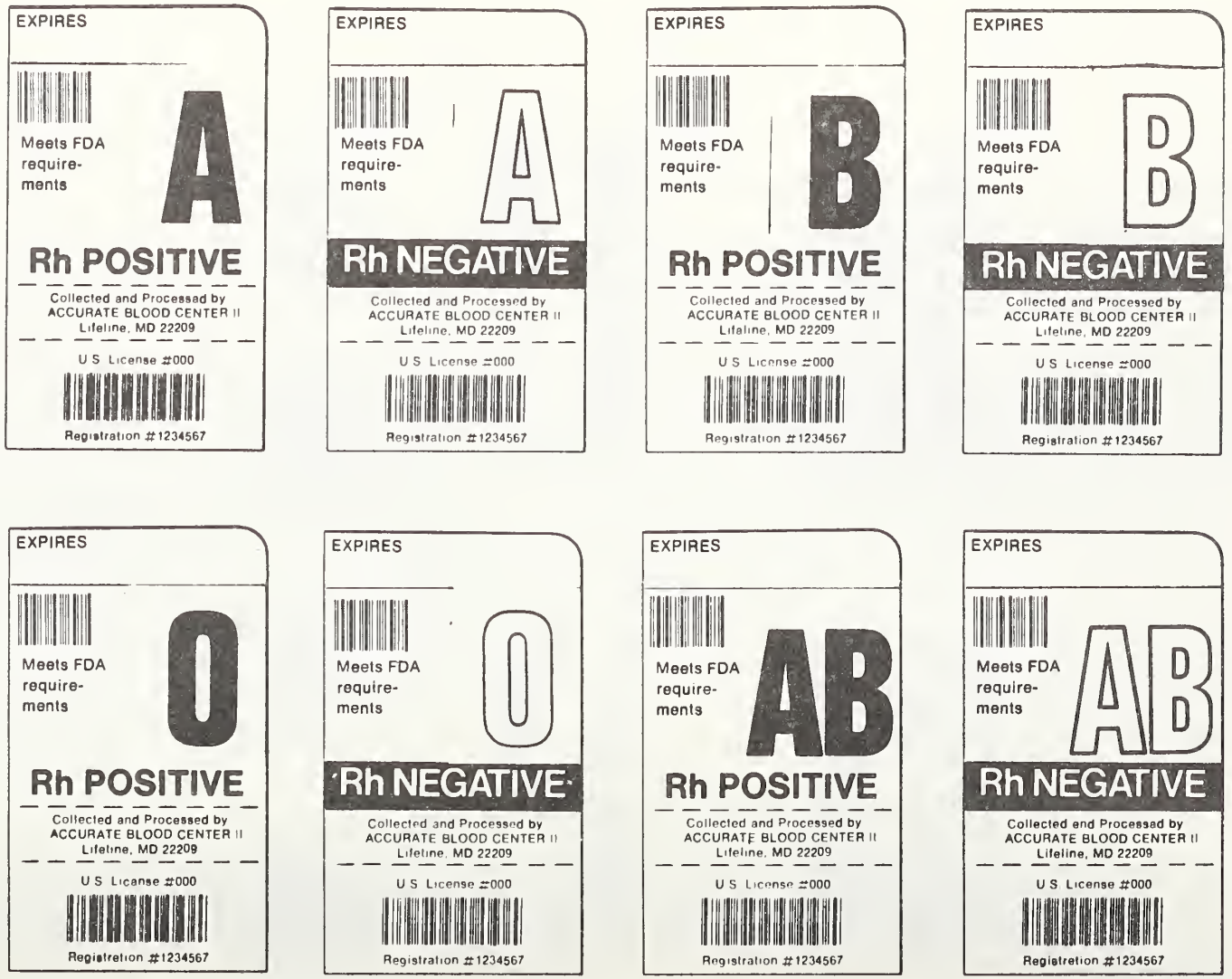


Figure 5. Testing and Grouping Labels

A Technical Profile of Representative Government Developed Data Element Dictionary/Directory Systems

Paul K. Brown

U. S. Civil Service Commission
Washington, D. C. 20415

There are many different perspectives on the use of data element dictionaries (DED)s to manage data resources. Those who intend to use DED technology to manage data resources can benefit from the experience of those who are already attempting to do so. An organization can evaluate its own data resource management needs and then compare them to those of other organizations with similar environments. It can then evaluate and select approaches which have already been proven in use. As long as requirements are first carefully defined, an intelligent choice can be made from among alternative ways of implementing a DED. These are whether to buy, to borrow, or to design the DED from the ground-up.

Key words: Data resource management; data element dictionary; data element directory; data element dictionary/directory; requirements definition; software selection.

1. Introduction

Because of my association with the FIPS Task Group 17 (TG 17) "Data Element Dictionary/Directories" survey of government DED's, I have been asked to share some of my observations with you. The survey has since been published as NBS Special Publication 500-16.

Well, over the past two years I've learned at least one thing. That is that there are as many conceptions of what data element dictionaries (DED)s should do as there are people using them. Everyone who can potentially benefit from a DED-systems analysts, forms and reports managers, data base administrators, computer programmers and others-have their own particular

viewpoint on data elements and data resource management. After examining the systems submitted to TG 17, one might expect that I could stand here and tell you which DED's do their jobs best or which are failures. Well I can't. Nor can I tell you which features of a DED are indispensable or which should be avoided.

Because of the multiplicity of tasks expected of a DED and the difference in the data management philosophies and needs of different organizations, TG 17 realized very early that the creation of a standard or recommended DED was an impractical goal. We felt that there were too many variables affecting the data management requirements of differing organizations which we could not adequately provide for in a standard.

However, since one of the original tasks outlined for TG 17 was a look at some actual live government DED systems, it was decided to use the survey to point out features of DED'S that others have found useful so that if an organization decided to implement a DED to help in managing data it wouldn't have to completely reinvent the wheel. Thus, the concept of a guideline replaced that of a standard.

2. Scope of the Survey

What is a DED? There was apparently a great deal of confusion on the part of those who initially responded to the survey as to what would be included by Task Group 17 in its survey of DED systems. Fifty percent of the systems described in the material initially sent to TG 17, although they were related to data management and/or systems design, were not considered DED's from an overall data resource mangement point of view. That is, they did not describe data elements and relate them to their use in the organizational environment. It is of interest to list some of the kinds of systems submitted to it that TG 17 chose not to include in the DED survey. These systems were excluded because they were not considered to be generalized data resource management tools of potential interest to those whose primary interest is the management of data resources. These included: data definition files for use with retrieval or report generation packages; bibliographic systems; and lists of elements whose primary purpose was to help in the design of a single system or application. Conceptual DED's and those which were in the planning stage were also excluded because we wanted the survey to reflect the "real world" and also, as we discovered, there was a disconcerting tendency for such planned systems never to materialize.

TG 17 recognized that various terms have been applied to DED type systems more or less interchangeably. Such terms as "dictionary" which implies definition, "directory" which implies location, "dictionary/directory" implying some combination of the two, and "catalogue" which connotes a list or description, have been, and continue to be used to describe, if not categorize, data management tools. Because no system we examined could be classified exclusively under any of the terms, TG 17 decided that for our purposes it would be better to think of these terms as capabilities of, rather than classes of, systems. These capabilities are present to one degree or another in most data management systems. The functions of a catalogue, dictionary and directory, particularly in more sophisticated DED's, are combined. Incidentally the task group quickly abandoned an effort to substitute the term, "data resource directory" (DRD), as a more precise term to describe a sophisticated DED. The details are just too gruesome to mention and we found ourselves adrift in a sea of semantics.

There seems to be an almost evolutionary process that leads from the implementation of one of these capabilities to another. There is and has been an instinctive tendency to first list, then define, and finally locate and describe in detail the uses to which data elements are put. It should be noted at this point that these capabilities of a DED don't have to be automated. In fact, TG 17 sees no problem whatever in using a well maintained manual system if it meets the needs of the implementor. The task group felt very strongly that we were not really trying to describe or compare software, but to describe total systems with certain common attributes.

3. Characteristics of Surveyed DED's

What are the informational components that TG 17 felt should be included in a DED? TG 17 identified eleven entities that can be relevant to an organization's use and management of data resources. Our look at the surveyed systems was in part a comparison against this model subset to see which of the entities had been considered and incorporated by implementors of DED systems.

Just to mention them, the entities identified by TG 17 are, in descending hierarchical sequence: Plans/Programs, Systems, Applications, Procedures, Files, Records, Data Elements, Reports, Forms, and finally, Documents. How does the scope of government developed systems compare to that described by TG 17? In general, as the experience of the implementor and the number of data elements managed increased, the DED systems tended to address more of the entities regarded as significant by TG 17. This seems to suggest that the greater the number of elements used within an organization, the greater the need to document what managers and technicians need to know about them.

However, the realization of exactly what information a particular organization really needs to know about its data seems to develop over time and often after some hands-on experience with a more or less unsophisticated DED. At least two respondents told TG 17 that trying to achieve sophistication in one great leap, especially where an organization's resources are limited, tends to bog down development and to hinder achievement of what should be principal objectives.

For those who like statistics, the average system surveyed addressed five of the relevant data management entities that TG 17 identified. Only one system addressed all eleven. Five of the systems had interactive query or update capability, and four had been implemented with data base management system software. One of them was a manual system. At least two systems have been significantly changed since NBS Special Publication 500-16 was printed. Four of the systems described have been transported to more than one installation.

4. Some Notes on DED Implementation

For those who decide that an automated DED is for them, the decision to make or buy can be crucial to successful implementation. I wish I could give you hard and fast rules about this decision but without knowing individual cases this would be impossible. There are both failures and successes for ground up in-house DED's, for generalized systems developed by commercial vendors, and for efforts to transfer DED's from one government agency to another. Two experiences illustrate the pitfalls involved. One user organization we talked to attempted to implement a commercial system, but had to abandon it and write a less ambitious in-house system because the powerful features offered by the selected commercial system got in the way of the basic capabilities that the user really needed. The designer of the commercial system had a different perspective of the problems of data control than the user had. On the other hand, an analyst associated with an effort to build an in-house DED indicated that if he had it to do over again he would buy, because his in-house effort tended to grow in scope until it exceeded practical limits.

I guess my own personal philosophy on DED implementation can be summed up in two points. First, it makes sense to buy if your requirements are well-defined before you sit down to talk to various salesman and if you can then find a ready-made system that fits your needs. Remember though, it is going to cost you enough in resources to do what you need without implementing bells and whistles and collecting data that you don't need. The effort to design a data management system should be no different from that involved in building any

other system. It should stem from careful requirements definition and sound functional statements. A commercial package or one obtained from another implementor can be a great cost and time saver if it is truly responsive to your organization's needs, and not just a solution looking for a problem. On the other hand, an in-house effort must have realistic and well defined objectives or it will wander around gathering irrelevancies until it falls of its own weight.

The second point of my personal philosophy is that I favor a "start small-gain experience" approach to data management. Set realistic initial goals so that you can make a successful step-by-step effort to identify, document, and control your data resources. I think it is true that most managers would rather see a small successful step into data management than a large failure. Collect only the metadata that you have a stated need for, not all that you think you may be able to eventually use.

5. Conclusion

Those starting out on the road to documenting and managing data use at the present time can draw on a wealth of experience-some impressive and some not so impressive-that was not available a few years ago. Draw as heavily as possible on this experience but remember, above all, to be careful that you are really solving your own problems and not problems relevant only to someone else's organization. If I had been asked 3 years ago what I would include in a DED, I would have given a very different answer than the one I would give today. I do know that it was helpful for me to read what literature was available and to look at existing systems for those features that seemed to me to be applicable to solving my own problems. The knowledge gained from my own experience and the experiences of those participating in this survey has at least broadened my perspectives on data management. I hope that what TG 17 has done will do the same for many of you.

Cargo Data Interchange System (CARDIS)

Robert J. Cavanaugh
Manager, Information Systems
International Sales Division
General Electric Company
New York, New York

Co-Chairman of CARDIS Standards Committee
National Committee on International Trade Documentation
New York, New York

Abstract

CARDIS is a system concept for a Cargo Data Interchange System to electronically exchange international trade data, being sponsored by the National Committee on International Trade Documentation. NCITD is a non-profit, privately financed, membership organization dedicated to simplifying and improving international trade documentation procedures, including information exchange by either paper or electronic methods.

Current methods of international trade information exchange is slow, with shipment documentation frequently arriving later than the cargo, thus delaying customs clearance and delivery to the ultimate consignee. Furthermore, preparation of paper documents is time consuming and very expensive because of the many documents used, and the repetitive use of the same data elements over and over again.

CARDIS will utilize modern technology including electronic communications transmitting data at high speeds, computer storing, sorting, retreiving, selecting data and formats, directing the transmission and the printing of required documents when and where needed. This will reduce delays now encountered at exchange points thus reducing costs involved with extra warehousing and handling expenses, clerical effort, errors and associated costs involved in current manual documentation systems in which the same data is entered many times for separately produced forms.

There will be three types of CARDIS centers to serve all industry and government needs. Type 1 centers will provide common user database facilities, store and forward message switching services and provide the mechanism by which the CARDIS network is tied together and operates as a coherent information handling system. Satellite type 2 centers will provide services to multiple users. This could be a major forwarder with a number of shipper clients or a service bureau servicing small forwarders and other shippers. Type 3 centers will provide direct links from a single major user who is able to provide his own computer facility and prefers to deal directly with the shipping information process. Subscribers to CARDIS will be required to use only those services which would complement their own systems and capabilities. Subscribers will be billed only for services used.

Text

Good morning ladies and gentlemen, I am Bob Cavanaugh, Manager of Information Systems for the International Sales Division of the General Electric Company. In addition, I serve as Co-Chairman of NCITD's CARDIS Standards Committee with my friend and colleague John Greene, Vice President of General Steamship Corporation, Ltd., San Francisco.

It is in the latter category that I speak to you today.

As Bob Porter has just informed you, the National Committee on International Trade Documentation is the industry supported organization dedicated to facilitating trade. NCITD is made up of many different individuals who represent 250 companies that are involved in our international trade. Bob and the Slide Show, "Linking You To Trade Profits", have amply covered the Who, What, When, Where and How NCITD has tamed the Paper Tiger over the past 10 years. My part in this program is to update you on the progress NCITD is making on the CARDIS project, and (for prospective) to give you an insight on how we at General Electric plan on making use of this Cargo Data Interchange System. Let's get started.

The technology involved in transporting our goods in containers has expanded faster than our ability to cope with the documentation problems. In fact, the efficient and speedy movement of the container was held up because of the paperwork problem that Bob discussed. Often ships sail and the cargo remains on the dock due to our inability to provide documentation on time and where needed. As a result of our study, we indicated that there were certain elements that had to be

reckoned with in order to attack this horrendous problem. We first had to:

- o Identify
- o Eliminate
- o Simplify
- o Standardize

We had to examine the practices, procedures, laws, etc. to develop the programs and methods to get at this very difficult problem area.

As a result of our indepth study, we developed programs, showed expected results, and developed a work plan to carry out our projects. An example of this was the problem involved between a shipper in one country and a customer in another overseas country. We questioned why it was necessary to process title shipping documents via so many individual parties in the banking community. The rapid movement of the freight indicates that they should be transmitted direct from the shipper to the consignee. As a result of this need our financial and insurance committee published the recommendations to importers and banks explaining how documentation could be simplified within the financial community.

The U.S. standard master, whether produced on a typewriter or on a computer, can provide many of the basic documents on a single one run system. Furthermore, through the efforts of our executive director, working through the Economic Commission for Europe in Geneva, our U.S. standard master was made compatible with the ECE layout key. This gave worldwide recognition and acceptance of this format.

In addition to the bill of lading we are now able to produce the dock receipt, drawback forms, shipper's export declaration, certificate of origin, insurance certificate, etc. It is safe to say that over 95% of all export shipments leaving the United States by vessel are shipped on an ocean bill of lading compatible with U.S. standard master. To further simplify this matter a special clausing was developed and now constitutes a shipper provided bill of lading. This eliminates the necessity for each ocean carrier providing his own verbiage on individual, separate forms.

NCITD's success with these and other programs explained by Bob Porter, make computerization a practicality. This was first brought about through the efforts of NCITD. In 1969 legislation was passed by the Congress of the United States which permitted the filing of the U.S. shipper's export declaration on magnetic tape, punch card, or paper tape on a monthly summary basis. (GE was one of the first to do this). That is what taming the paper tiger was all about.

Our efforts have, by no means, been limited to our own shores. Our executive director makes quarterly trips as the industry representative to the U.S. delegation to the meetings of the ECE, the United Nations organization charged with international trade facilitation. Furthermore, there have been many sister organizations like NCITD

formed throughout the world. Having arranged the data in recognized standard positions has greatly reduced the language barriers of the past. All are adopting NCITD's approach to streamlining international trade documentation through elimination, simplification, standardization and computerization. These elements have also been linked together by our trading partners throughout the world. We can report that much has been accomplished these past 10 years in reducing this paperwork burden. Now to relate this to our own environment at General Electric.

Long ago we recognized that the documents involved in our export business were much more complicated than our domestic invoices for the same products. The complex interrelationships of various organizations and functions performed within the International Sales Division required considerable effort evaluating and reevaluating our business systems. Through the efforts of the past 20 years, we have been able to bring order to this interrelationship through our computerized export documentation system called the export system. In 1958 we started computerizing our export documentation. We are now working on our fourth generation.

The export services organization acts as an interface between our domestic product departments, the overseas customer and our International Sales Division. Our orders range from simple AC motors to major turnkey power plants, with sales in practically every country of the world. Export Services is mainly concerned with the customer's order being properly placed at our General Electric product departments or outside vendors. They then arrange shipment of the material and supply the data for preparation of the necessary export documents, which in turn are distributed in accordance with the customer's instructions. The new system must parallel the new needs of our business which have changed substantially in General Electric both internally and externally. Furthermore, the technology is now available to overcome our major limitations.

Our new system will provide remote printing of documents, remote on-line entry, remote inquiry. Terminals will be located at the product departments. Our people have come up with an acronym for the new export system, which spells ACCESS. ACCESS means Automated and Comprehensive Customer Export Service System. ACCESS will be communications oriented, responsive to user needs, abreast of the State-of-the-Art and be flexible to meet various user needs - but by no means all. It cannot do the whole job for GE by itself. To better understand why not, let's take a look at the problem.

Today we have a transportation system which has the capability, notwithstanding the legal and governmental restrictions, of being very efficient with containerization, intermodal shipping, air cargo, etc. This transportation network goes international also. The problem we face is that this highly effective infrastructure for our transcontinental and international movement of freight has matured over many years. However with today's worldwide economy, the question arises whether we can allow the same maturing cycle to happen in our trade documentation development. It seems we should be using modern technology like computers, to accelerate this development. In particular each of us in doing our share of trade documentation should have an overall long term plan.

To be cost effective we must mutually exchange data whenever possible thus minimizing the most error prone and costly re-entry of data throughout the cycle. In addition each of us in our own systems has to look continually for the best possible cost benefit ratio. In many cases we might be able to find significant value added services that could be handled much better on a pooled or shared basis.

Today our computer systems output paper. Lots and lots of paper. This is the most inefficient and ineffective way to use computers. That is, like large, old fashioned printing presses. In fact, if we looked at today's integrated computer systems you will find that the method of integration, in more cases than not, is paper. Undoubtedly there are some of you, like us in General Electric, who are doing a limited amount of data communications. However, it is generally still on a very selected and point-to-point basis. CARDIS or the Cargo Data Interchange System being proposed must parallel and work in conjunction with our cargo handling systems. To do this effectively we have to build upon our physical distribution systems and integrate our data manipulation systems with them. This means they must be:

- o Capable of decentralized operation, and yet,
- o Highly integrated.

In addition, we must allow the normal marketplace options that are very common to us in the physical distribution area. That is:

- o Looking for the best return on investment,
- o Looking for the best services, and
- o Having the ability to use multiple vendors.

We must in effect establish our Cargo Data Interchange in an evolutionary manner, plus insure it is integrated with our transportation worldwide scheme. Our information flow cannot tie up our transportation network as it does today.

It has always been NCITD's objective to build upon the paper handling improvements and achieve computerization. This overall project was given the acronoym CARDIS. CARDIS stands for Cargo Data Interchange System. CARDIS is the means to apply U.S. business systems computer technology and experience to the jungle of trade documentation. It keys itself upon the entering of data only once and the electronic data interchange of all required data to those who need it throughout the entire cycle of international trade. CARDIS has ambitious goals and many problems to be resolved, since international trade is intertwined with many countries national rules and regulations.

CARDIS has many features. One major requirement is to interface with other systems, such as in-house applications like our ACCESS system and overseas systems like the LACES system at London's Heathrow Airport. Another major feature was the requirement to provide the capability for value added services. These services could extend from pooled transoceanic data communication networks to currency conversion aids. In summary, CARDIS was the "final" link in NCITD's chain for attacking the trade documentation problem.

Before we go sailing off into the wild blue yonder on the theory, it might be a good idea to take a look at the practical capabilities of such a system concept by looking at a demonstration. This demonstration is possible, because of the work that has already been done at producing aligned forms in the paper world. This particular demonstration would be output oriented. A key to the output will be the U.S. standard master as developed by NCITD. This particular format has multiple purpose capability. Our demonstration will highlight the building of information by data entered only once. Appropriate data interchange via computers will produce the necessary paperwork required to move the material by our transportation system.

We will start this demonstration at the beginning. Most of the data originates with the shipper preparing the domestic inland bill of lading. We will follow material movement by various supporting documents, plus the commercial invoice. The commercial invoice can be aligned with the U.S. standard master, or the new floating field concept, which is much more suitable for computerization. The aligned bill of lading can be produced from the data supplied by the various parties. Because of time limitations, there are many action documents that we will not demonstrate tomorrow morning. It should be clearly understood that they can be produced not only when needed but where needed. However, the objective remains to eliminate the need of paperwork. This is particularly true for the U.S. shipper's export declaration and the notification of exportation for duty drawback. Included in this demonstration will be the retention of the data thus allowing for the production of the ship's manifest.

The input to this demonstration will be terminal oriented in conversational mode to illustrate low volume data entry. Data transmission with existing business systems will be the normal way for data interchange when the parties involved are computerized. The demonstration will use coding to speed up the input of the data and to highlight the capabilities of some of the value added services. These can be performed by various CARDIS rated centers, by in-house business computer systems, or by service bureaus. The demonstration will also highlight some of the tracing features that the system will be capable of providing. We will conduct the demonstration on an informal basis, this afternoon from 2:00 until 5:15 p.m. and tomorrow morning from 9:30 a.m. to 1:00 p.m.

Until recently, CARDIS could have been thought of as another good old U.S., long term, simplistic answer to a global problem. But CARDIS research was started a couple of years ago under the U.S. Department of Transportation sponsorship. In the United States a study to show the concept was technically feasible in total, left the impression that one very large physical computer system was being planned. (The bull really flew then). Also many legal and security considerations were not being resolved as fast as computers and systems were being planned. Finally the idea of satellites and worldwide telecommunication networks started to raise some feasibility questions in people's minds. Not technically, but economically. Because the dollars were going up seemingly very fast and the savings were not being clearly defined. Also, internationally, meetings with the Economic Commission for Europe and other trading partners somehow

conveyed the immediacy and completeness of our research. To that extent U.S. CARDIS was pictured to be a problem -- and a very near term problem at that. All in all, the CARDIS programs had many problems. Thus a halt to the headlong charge was inevitable. It is NCITD's plan that order will come out of recasting the program to a more durable and logical cycle.

Let us look at this plan from the standards point of view. CARDIS must serve the needs of both business and government. In order to ensure this we need an overall system concept. This overall concept will also ensure the integration of each of the parties in the NCITD chain to break the documentation bottleneck. CARDIS should, and in fact, must work with our in-house systems in a positive and supportive role. The standards of a CARDIS concept should increase the value of automated systems within the served businesses. The goal of CARDIS is to keep the production of documents to the minimum number that are necessary to support the legal and governmental requirements for international trade. The standards are being developed to allow an evolutionally introduction of the CARDIS concept into our business information systems. The key to any standards activities starts at the basic data element level. Additional standards work must be done on the computerized versions of required documents. Foreign requirements for data interchange must be actively and constructively factored into our U.S. standards. We must ensure data compatibility and interchangeability between the various parties in the international trade transaction. For the forseeable future, we must integrate with the paperwork currently required. CARDIS is our long range plan to make it work for us in the United States. We need each of your company's cooperation and active participation. There is a lot of work yet to be done if we are to succeed.

In total here is the CARDIS approach towards the development and issuance of standards. First of all you must define all the trading requirements at the document level so that it can be tied into today's requirements. Data elements necessary to produce the document on a computer must be identified. Finally, we must provide the procedures for the data interchange of those elements. Once the computerized procedures and documents are developed, they must be continually coordinated internationally to ensure the long term life cycle of the CARDIS concept. Meanwhile we intend to integrate with the Transportation Data Coordinating Committee the data elements and interchange requirements of the CARDIS concept. We want these requirements to be factored into their EDI protocol. In a parallel procedure, we will be integrating with ANSI and the ISO via the National Bureau of Standards. In addition, the CARDIS standards will definitely support the concept of a multi level decentralized CARDIS centers, including qualified in-house systems. The key here is not a large system and not one data base. Data should be entered only once and made available under a distributed data base concept according to the legal, security and company requirements of the parties involved. Finally we will support this development of the value-added CARDIS centers. Some of the value added services that we see include trade reference libraries, associated master files and related

services as indicated. In addition we look for a value added service of cost effective international data communications. And finally we look for "pooled services" for the economic processing of the needs of small shippers and freight forwarders for the cost effective preparation of the required trade documents.

In the past I am sure each of you continually tried to benefit from NCITD's activities. GE has. I think you can see where computerization by an evolutionary long term plan can save your company significant overhead expenses. Just look at your uncompetitive bids, your capital lockup in blocked freight, receivables or inventory. NCITD needs you in this, as in its other activities. Together, we can link your company to improve trade by the effective use of computerization -- and help our national goals at the same time.

In conclusion, most U.S. industry and our government are quite concerned about the direction and speed of the U.S. economic growth. G.E. is, of course, quite vocal in this area. Dr. Thomas A. Vanderslice who is our Vice President and Group Executive of many of our high technology industries, including Information Services Business Division (ISBD), has been highlighting to various audiences the key that technological innovation has to play in the growth of the U.S. economy. In particular, he has recently stressed the fundamental relationship of technology, including computer technology, in relationship to increased productivity. This is especially true in the highly competitive worldwide markets where so much of the growth in the world's gross national product is expected.

To see the value of technology, let us look back at the statistics of the period 1950 through 1974. This is a period where we had seemingly unsurmountable technological leadership, with corresponding advantages in our products and processes. If we compare the U.S. industries according to their technological content, we first find that the high technology base industries grew almost three times faster. The growth in productivity, or the output per employee in these high technology industries was doubled. Conversely, the price records reflected a 6 to 1 ratio of the higher effects that inflation had on the low technology industries. An even more significant finding was that gains in output per employee was not at the expense of employment. In fact, employment in high technology industries grew almost 9 times faster. Finally, the same kind of favorable ratios prevailed in the terms of international trade with a positive contribution to our balance of payment situation.

There have been changes since 1974 and one of the most significant is the elimination of most of our technological advantages. This is in no small part due to the suffering of malnutrition within the research and development activities within the United States. There is a definite economic side of technology. Together, industry and government must look for ways of applying technological advantage to today's problems, growth productivity, inflation, employment and balance of payment.

The incentive is here for applying one of our major technological advantages, the computer, to the very inefficient trade documentation practices we currently face. Time is of the essence. GE is bullish on exports. But we are concerned because ships and planes don't move international freight, paper does. Together we can change that.

I thank you for your interest and I hope for your active support.

American National Standards Institute (ANSI)

Ward Duel*

The American National Standards Institute (ANSI) was founded in 1918, at a time when standardization in this country was in a chaotic state and international standards were only a dream. There were five founding professional societies and three government agencies. The present base includes 180 professional societies, 1000 companies, and local, state, and federal governmental agencies. ANSI is diverse enough to allow each to have his own perspective of what it does. Let me give you mine.

ANSI does the following:

1. Identifies needed standards and sets a priority for their completion.

2. Proposes standard development work to be done by competent organizations.

3. Assures that everyone's interest, including consumers, is protected.

4. Supplies standards' writing organizations with effective procedures and management services, to ensure efficient use of manpower and financial resources.

5. Ensures that needed standards are developed on time.

6. Accepts voluntary consensus standards that:

   1) Express desirable technical concepts.

   2) Establish a general frame of reference for safety and effectiveness.

   3) Speak to three levels of compliance. These are:

      a. The "may" level, which refers to a desirable goal which might not even be economically or technically achievable when written.

      b. The "should" level, which implies a currently achievable goal which ought to be accomplished.

      c. The "shall" level, which is a mandatory term that usually refers to a minimum standard.

Clearly, "consensus standards" are not specifications or minimum regulations.

Before a standard can be considered for approval, it must be derived by competent national leadership. The "American National Standards Institute Procedures for Management and Coordination of American National Standards" outlines the requirements for developing standards. Simplified, there are three methods that can be used.

1. The Accredited Organization Method.

2. The American National Standards Committee Method.

3. The Canvass Method.

The Accredited Organization Method

This is the method that I, personally, am most acquainted with, because I have worked with the American Society of Mechanical Engineer's task force on standards for safe and sanitary design of food, drug and beverage manufacturing equipment (ASME-FDBME)

Any organization substantially involved in standards and willing to comply with the criteria listed in the prior mentioned publication may apply to and be accredited by the Institute. In developing standards, the organization must provide an opportunity for participation to all concerned national interests, must maintain a balanced committee, must consider and answer all objections, and must report to a supervisory body, who in turn, must attest to the fact that all procedures have been followed and conditions met.

Accredited organizations' committees

---

*Mr. Duel is Assistant Director, Department of Environmental, Public and Occupational Health, American Medical Association, 535 North Dearborn Street, Chicago, Illinois 60610. He is a member of the American National Standards Institute Medical Devices Standards Management Board.

are required to achieve consensus through a system of balloting in which all negative votes must be accompanied by a stated reason. The committee must then make an effort to resolve each dissenting vote. Sometimes compromise solutions must be accepted which may not satisfy all the objectors, but the newly worked out solutions must fall within parameters that all can accept.

There must be balance among the interests of committeemen, and no one class of committee members can be in the majority at any meeting where decisions are made. In the ASME-FDBME Committee, the classes are professional sanitarian, manufacturer, user, professional society member, and government and academic representative. Three levels of government are represented.

I will not lead you all the way up the tortuous trail of developing a standard, but I would like to mention that when submitted by an accredited organization to ANSI, each standard is considered as only a "Proposed American National Standard". Periodically, ANSI publishes the list of proposed standards and solicits public comment. Only after questions raised by the public are addressed, and it is determined that a national consensus exists, does the "proposed" standard be an "accredited" standard, i.e. an American National Standard. Let me assure you, the public respondents always include several well informed experts, who often disagree with some aspect of the proposed standard, or who feel that the new standard will hurt their business. Those objecting to the final decision have recourse to several stages of appeal, which time will not permit me to enumerate.

The American National Standards Committee Method

This committee method is only activated when it is determined that standards in a specific area are needed and will be used, but that no existing standards' writing group has them currently under consideration or intends to study them in the future. The appropriate Standards Management Board assigns a secretariat to do the work. A secretariat is an organization, not an individual, which is responsible for administering the business of the committee according to established ANSI procedures. There is, on occasion, a bit of competition over the secretariat assignment. There are many procedures to follow that assure all points of view are considered, and all individuals and organizations are given a chance to participate, but when the process is complete, workable standards emerge.

The Canvass Method

The canvass method is used when an organization believes its standard should be adopted as an American National Standard. The proponent must be willing to submit evidence that overall agreement on the acceptability of its standard exists. Most of the ANSI rules applicable to this method relate to who should be included on the broad-based canvass list, how should objections be handled, how should the canvass results be reported, and what action is necessary for acceptance by the Institute.

I am a member of the ANSI Medical Devices Standards Management Board, and at the same time also liaison to the FDA for the American Medical Association. Part of my latter role includes studying the possible effects of the implementation of the Medical Devices Amendments of 1976, P.L. 94-295. So, in a sense, I wear two hats. There may be a problem of conflicting purposes between the standards prepared and used by government regulatory agencies, and those prepared by standard writing organizations. Using the example of medical devices, the FDA is authorized to promulgate compulsory regulatory standards that will insure safety and effectiveness, while ANSI standards frequently contain some idealistic sections, and compliance is intended to be voluntary.

Recently, there have been moves to withdraw ANSI standards, because the government feels chaos will result if idealistically drawn, voluntary, consensus standards are used alone to regulate industry. One of my hats tells me that it is ridiculous to restudy and redo standards that were hammered out in hard sessions by a broad representation of interests. The question then arises, what then should be done?

It is my personal belief that 1) the government should not interfere in areas where good, voluntary standards exist, and where the trade is competently regulating itself. 2) The American National Standards Institute should identify minimum requirements of safety and efficacy, and delineate clearly the may, should, and shall levels of achievement. If ANSI did this, its standards would be more suitable for governmental use. 3) Where there are no

standards, or where standards are inadequate the government should use the existing, voluntary standard setting system and its procedures. To make their standards acceptable for government regulations, desirable goals which are not achievable and achievable goals which are not essential, should be identified as such. 4) Lastly, government and voluntary standards groups must agree on the rationale and scope of the proposed standard before the developmental stage can begin.

I came here with an assignment, to introduce you to the American National Standards Institute system for producing standards and with a personal mission. The above completes my assignment. My mission is to encourage you as data processing people to carefully manage data that comes to you, upon which decisions of health are resting. The data which is derived from many sources will some day point the way to causes of cancer. It will show which environmental pollutants are dangerous to health and at what concentrations they are harmful. Standards can then be reevaluated and causes eliminated. The data will pinpoint potential victims.

I mentioned previously that I am currently working with other committee members on the implementation of the new medical devices law. To carry out this assignment effectively, we must be able to assess vast amounts of data. The committee and governmental representatives must compare the accuracy of different existing techniques, and declare less accurate techniques obsolete. Laboratory directors will be comparing their laboratories with other laboratories in order to review their effectiveness and improve weak spots. Much will depend upon the data that is presented.

I haven't even touched on other vast areas that are affected by medical data: clinical trials for drugs and devices, medical insurance costs, medical procedures analysis, and so on. Much is dependent upon you.

Management of Data Elements in Information Processing (Needs for Standards in the Health Care Field and a Sampling of Current Applications)

Gerald J. Duffy
Vice President
EDP and Telecommunications Services
Blue Cross Association
Chicago, Illinois

The term "standards" means many different things to different people. The objective of a particular "standard" may be to contain costs, to simplify a work effort, to control the use of resources, or combinations of these and other items. And the health care field is no exception in this regard. People feel threatened or inhibited by standards; others feel more secure with standards; and attitudes such as this affect productivity one way or the other.

And where a standard is "placed" or "defined" is important in health care, as it is in any business environment.

Congress, the American public, including the providers of health care and health insurers, all want to contain costs of health care and increase performance within the health care delivery system.

I'd like to explore for a few minutes how standards can aid in that direction, what kind of standards are often used, whether they affect a local or more general environment, at what point in a business process they might be "placed," and some of the dangers in applying them in the wrong place in a process or at the wrong point in time. Within the health care field, in particular when we process a claim for payment of health care coverage, we define standards for the level of care a patient receives and the length of stay for a patient with a specific diagnosis. And these standards must be established so as not to pervert an important principle of health care, namely, what are the appropriate procedures for a given disease. In other words, the standard must be sensitive to the delicate balance between appropriate health care and the need not to overutilize costly health care resources.

And many in the health care prepayment and insurance field set standards for productivity, turnaround on claims payments, claims inquiries, etc.

These standards can be set to accomplish a key objective such as better service or cost containment, but such a standard must be well-defined, realistic, and subject to periodic review, over short enough time intervals, to allow adjustments so as to avoid bad side effects.

Consider a health insurance program over a large population, such as some of the national health insurance legislation proposed over the last five years, much of which has included the concept of a national health care credit card. If such legislation is passed, or if a health care credit card becomes popularized by other means, perhaps standards may be appropriate for the computer terminal needed in a doctor's office to verify a patient's

eligibility and benefits remotely. The terminal in such a situation, we must realize, must be simple for any receptionist or medical aide to use, regardless of the size of the office, the rate of attrition of employees using it, and so forth. It must be inexpensive both in initial cost and operational cost. It must be capable of reading patient information from a card, such as patient ID, any deductibles or copayment levels, and other parameters required by the health care program. And the terminal must be capable of alerting the terminal operator regarding the patient's eligibility status, among other status elements required by the program. This need for a simple, low cost terminal seems more rational when one considers that such terminals serving similar populations for other services, such as super market retailing, have often been too complicated for an untrained employee to operate or for a customer to operate. So also, vendors of terminals have made sincere but generally unsuccessful attempts to standardize terminals resident in hospital admissions offices for years, partly because the terminal had been initally designed for a general computer application market and not for its customer in this case, namely, the hospital administration market. And terminal vendors have suffered similar growing and marketing pains in the banking industry where today's terminals are becoming more standardized and even usable by the general public for deposits, withdrawals, and other transactions.

In such an example as a health care credit card terminal, the standard may indeed be a very simple one, namely describing a durable, low cost device which has a reader mechanism always in the same location and usable in one way only, with a light and/or audible alarm which signals the status of the card bearer for a very small number of parameters, that is two or three perhaps. One may quibble with this and call these "requirements" rather than a "standard," but these terminal requirements become a standard if they are used widely, and there are many economic pressures which can translate such basic requirements into a "de facto" standard.

I would say that such a standard has a reasonable expectation for success because it serves a large but very simple application where the invoking of the standard constitutes little or no threat to anyone save perhaps, the specialized terminal operator of yesterday who, with such a simple terminal, would look like the legendary fireman on a diesel locomotive.

Health care provider terminals need some standardization to save cost, and some future applications may precipitate some standards in this area.

And, as long as we're looking at the computer terminal, which after all is very close to the end product business user in the doctor's office, in the hospital or in a claims processing office, let's follow the trail back to the host computer which processes that subscriber eligibility check, the subsequent claim and the payment of benefits. In that host computer we find a mammoth system which performs all of these functions and more. Within the Blue Cross and Blue Shield System we have a development underway that illustrates a high degree of standardization of software by using a method of decision table processing to automatically adjudicate a claim. And this software system may be resident on any one of a number of different IBM computers with different hardware configurations and operating systems. Thus we have a standard way of defining benefits and the automated adjudication process which uses any benefit definition as a processing algorithm.

Now let's look at the communications conduit which links computers.

In the health care field, the hardware and software available from

computer vendors is different especially within the context of the vendor's network architecture. Perhaps truly standard network architectures can be defined for general classifications of telecommunications network needs and vendors can make these architectures available through their network processing and terminal product lines. Then the competition would be for the best and cheapest way to carry out that architecture. Such standards would no doubt need revisions at least every ten years to avoid the perpetuation of obsolete technology but could help significantly in lowering the administrative cost of health care.

When computer vendors are approached today with the question of a standard for a language such as COBOL, some respond by talking about standardization among vendors, of the hardware components most used to translate a COBOL program into machine code. Vendors now seem to question the need for a completely standardized COBOL language by emphasizing a standardization of the hardware which produces the language translation instead.

So here we have a question of the place at which a standard my buy us the most within a process. Placing the standard in either place can reduce the cost of health care.

Finally, perhaps some standards are relevant in the business area which regulates most of our costs, namely, "management." One management consulting organization has created a successful business by showing that six components of management "climate" correlate heavily with the productivity of a management unit. And many of its customers have created a more meaningful management "climate" by accepting the definitions of those six components as a local management "standard" which precipitates discussion of internal management problems from the same definition base for all managers.

In summary, maybe we need more standards, but perhaps we are applying them in the wrong areas, in the wrong place and at the wrong time.

KEYNOTE ADDRESS OF

PETER J. FINNERTY
VICE PRESIDENT
SEA-LAND SERVICE, INC.

THE NEED FOR STANDARDS IN DATA TRANSMISSION

SEPTEMBER 29, 1977

Good morning and welcome to Washington, D.C. A fitting location for this morning's convocation on International Trade Data Standards. Washington is the world capitol for data from stuffy reports to slick computers, from red tapes to telephone taps. Seriously though, not only does our government monitor our foreign trade in the world, but this bit of geography on the Potomac has witnessed some remarkable upheavals in foreign commerce caused by the march of American technology.

Such opening lines, along with the technical headings on the agenda, probably lead you to expect me to launch into recitation on data bits and facilitation hits as you digest your breakfast. Well, with apologies to any died-in-the-wool technocrats in the audience, I will leave the specifics of the how to the experts that follow. Instead, my opening discussion will communicate the why.

The dimensions of the opportunity that lies before us in International Trade Data Standards are, believe it or not, nothing short of exciting. It comes as no surprise to a group of specialists such as yourselves, that the enormous volume of commerce between the United States and other nations is currently a priority issue second to none. The implications reach into the lives of every citizen, touch upon the programs of about every agency, directly impact our national economy and are especially critical to the huge number of companies involved, from mom and pop operations to multinational giants.

From petrodollars to comparative advantage, from floating exchange rates and orderly marketing agreements to technology transfer, the all-transcending "International Order" of the present cource of events is nevertheless, merely a chapter, in a much more lengthy text.

Charlie Hiltzheimer, our Chairman of the Board and principal mover and shaker at Sea-Land, has several valuable nuggets of wisdom, one of which is apt for our discussion: "see it big, keep it simple." Obviously, the continuing struggle by nations and peoples for superiority in the intensely competitive marketplace of world trade is a subject of global proportion. However, discerning the key pieces in the overall scheme of trade is too often obscured by a wave of explanations in complex terminologies. If we aren't careful, such a litany of "why nots" will confound the best of us.

It was Adam Smith, in his wealth of nations, who noted that sophistication and complication is too often a means of evading a simple truth. In our case of International Trade Data, I submit that we stand at the threshold of another small, but important step for mankind. The simple truth involved is that despite the myriad obstacles and rather forbidding multiplicity of institutions, companies, agencies, commodities, etc., Foreign Trade Data will be standardized and brought into the age of modern information technology. That may sound like a squeak to many of you, rather than a roar. You may think if he considers that exciting, he probably gets pepped up over making out his tax return too. Well, the fact is that even the I.R.S. has recognized the compelling logic of using the computer and manages information in a manner that ought to prod us into action.

The I.R.S. knows what successful enterprise the world over has discovered, extraordinary achievement begins with information. As at least one computer company has pointed out, make information easily and instantly accessible and you set your energies free to explore the upper limits of the possible. Look around us and you will see many supposedly impossible achievements that involved prior agreement on standards of some sort. Electronic banking, grocery store scanners, automated airplane tickets and hotel reservations and even international direct dial telephone calls.

The simple fact is that the goal of Trade Data Standards will survive the gauntlet of human disagreement because the economics and management motivations of the many functional sectors involved in international trade demand it. Add to that the personal dedication of the many experts that have spent over a decade reshaping the trade information portion of the overall Foreign Commerce picture and you'll see that the necessary elements exist today to make this goal happen. The key point is to recognize that we do have a hard core of common agreement surrounded by a rich variety of individual differences. Beyond that, it will require ample amounts of patience and persistence on the part of the participants. This latter factor is at least as crucial as the balance of the equation.

What do I mean by those terms patience and persistence? Not simply one good try and if it doesn't work you'll pick up your marbles and run home. Not at all, because ultimately this endeavor will, like many standards efforts before, reach far beyond the shores of our great nation and involve numerous foreign interests, many of them hostile to our position for other than fair and equitable reasons.

The scale of commitment that this task requires can be compared to the spirit exhibited by another American. At the risk of sounding trite, his experience is worth citing. That individual failed in business twice, was defeated for local public office three times, defeated for the Vice Presidency and finally, twenty-nine years later, Abraham Lincoln was elected President in 1860. That is not, of course, to suggest that we'll have the luxury of three decades to sort out our differences. No where near that long, gentlemen.

Modern communication and transportation have compressed the periods of reaction time or advance notice, if you will during which we either meet the test or pass the baton of leadership to the people or nations that possess the will to succeed in this endeavor. After all, where is it written that Uncle Sam and we Americans are entitled to remain "king of the mountain" forever? Not unless we work at it and hard!

I'll wager that most of you will attribute the phase, "survival of the fittest" to Darwin and his thesis of evolution. In fact, that term was coined by Herbert Spencer around the turn of the century in his writings on economics. Of late, many other American interests are struggling with the hard reality of world trade and the economics of comparative advantage, both artificial and natural. Even container transportation, which has achieved unprecedented increases in productivity from American technology ranging up to levels of 400 percent must cope with such factors as state-owned carrier competition, primarily the Soviet Union's, which throws the profit motive and traditional market motivations right out the porthole.

This year's International Economic report of the President sent to Congress in early January estimated that during 1976 the U.S. Merchandise Trade balance suffered a deficit of 13.8 billion dollars, of which, by the way, 8.7 billion represented the cost of freight and insurance on goods transported. That is, sixty-three percent of our merchandise trade deficit for 1976 represented payments, freight and insurance payments, to foreign carriers.

Projections by the Carter Administration in the last few days indicate that the trade deficit in 1977 will be between 25 - 30 billion dollars, the largest in America's history. Commerce Secretary Juanita Kreps said in Tokyo on Tuesday that the "Major hazard we face in having such a large deficit in not a hazard to the strength of our economy, but rather it lies in the fact that we may incur a strong sentiment for protectionism."

I cite the deficit figures and the risk of trade problems because our subject related directly to the solution to those problems. The solution is to encourage the flow of trade, not impede it. To that end, I want to both thank and congratulate the National Bureau of Standards for fostering the International Trade Data Standards effort. Now that the initial momentum has begun, it is vital that the appropriate follow-up take place.

Data standards represent that critical opening step in upgrading World Trade Data Transmission. The momentum is building for a shake out in foreign trade that could rival the turmoil seen in the securities industry as modern information technology is applied. Properly accomplished though, the Trade Data System will serve all factions and not just entities that push for concentration and economics of scale.

For our part in the private sector, I think I see the initial signs of natural economic pressures coming to focus on this particular problem. Foreign Trade Data communication is increasingly the weak link in the chain. World trade interests here and abroad will press harder and harder to correct that situation and once the initial but significant challenge of standards has been overcome, there's no way to estimate the extent of commercial and social benefit that could be unleashed. Quantum jumps in productive activity, conversion of now wasted effort into value, satisfaction of need and relief of want are reasonable expectations in this yet unfolding human experience. I do think that's an exciting prospect.

I am even more confident now than in the Fall of 1966 when I had the good fortune to begin working with many of you to improve the international competitiveness of our companies, our industries and our country that we have the right plan and a winning team in this contest. It's right there for the taking. In not too much time, it will be clear to all of us and everyone else whether we've met the challenge. Let's see when that international appraisal takes place, the verdict on the U.S. system is, in the words of the latest James Bond tale, that "Nobody does it better."

Data Processing in Natural History Museums

T. Gary Gautier, Stanwyn G. Shetler[1]

National Museum of Natural History
Smithsonian Institution
Washington, D.C. 20560

and

Craig Black[2]

Carnegie Museum of Natural History
Pittsburgh, Pennsylvania 15213

Besides being places for relaxation and entertainment, museums also preserve objects that tell us about the biological and physical world in which we live and about our culture and history. Museum specimens are like books in a library in that they are the reference materials for cultural and scientific research. Hundreds of millions of specimens are contained in the over 6,000 museums of all kinds in North America. A few of the larger natural history museums alone contain tens of millions of specimens each, which poses a major problem for their information systems.

Recorded data about objects are almost as valuable as the objects themselves because they provide essential information which cannot be determined by examining the specimens. Complete conventional museum record systems consist of field notes, accession papers, permanent labels, catalogs and collection indexes, but many museums can afford to maintain only minimal documentation, which makes their collections more difficult to use. Museums are turning to data processing to improve their collection documentation. Several software packages have been especially designed for museum use. Data bases are being built for newly collected specimens as well as for the massive existing collections. The benefits in terms of greatly increased collection indexing and new uses for museums data are much greater than in conventional systems.

Intermuseum data exchanges are a goal in the museum community. Scientific disciplines and museum organizations form major channels for coordination of standards and

[1]Chief, ADP Program, and Associate Curator of Botany, respectively.
[2]Director

procedures between natural history museums. Data standards are already similar between natural history museums, but experiments with data exchanges are needed to discover where greater coordination may be needed.

Key words: Collections, data elements, data problems, data standardization, museums, natural history, networks

1. Introduction

Effort is being made in the museum community to upgrade the quality of museums and to improve and build upon the many services they perform. A major part of this activity involves automated information systems, since museums are basically information centers. The purpose of this paper is to briefly introduce the non-museum person to what museums really do and to the current status and problems of their efforts to implement information systems and data standards. The paper emphasizes natural history museums because the backrounds of the authors are in Paleontology (Gautier and Black) and Botany (Shetler). Other kinds of museums are making great strides in information processing, too, and they face other kinds of problems.

The following references will provide good reading for those who want to learn more about museum data processing: Chenhall [2], Shetler [5], Squires [6], and Vance [7]. Twelve other good papers are available from the Museum Data Bank Committee[1].

2. Overview of Museums

Museums are generally viewed by the public as lovely, aesthetic places which provide relaxation, entertainment and a general kind of education, but museums are a great deal more than that. Museums preserve for us and for future generations the objects that tell us about our past and our present, about the biological and the physical world in which we live, how the world came into being, and about our culture and our history. They are also centers of study for a full range of scholars from the high school student preparing a term paper to the advanced historian or scientist looking for answers to more subtle problems, and they provide a broad range of information and identification services to law enforcement officials, government agencies, businesses, educational institutions and to the general public. In the way they work and in the things they do, museums are similar to libraries - they hold reference materials - the objects - which are made available for study, and which must be catalogued, indexed, and stored in a systematic and careful manner, and they are staffed by highly trained professionals.

The Official Directory of Museums [1] lists over 6,000 museums in the United States and Canada, and most of these fall into four major categories based on the kinds of objects they hold and display - natural history museums, art museums, history museums, and science and technology museums. Most museums specialize even further. A museum may show a subject-matter specialization, eg. African art, plants , farm implements, or American Indian objects. Some museums emphasize public educational programs and exhibits, whereas others give great weight to building representative collections for research or to preserve particular aspects of culture,

[1]Address: Dr. Robert G. Chenhall, Chairman, Museum Data Bank Committee, Margaret Woodbury Strong Museum, 700 Allen Creek Road, Rochester, N.Y. 14618

nature, or science. Museums also specialize geographically, some emphasizing local or regional subjects where others may be of national or international scope.

The specialization pattern of a museum depends on many factors such as its age, the interests of its staff, the manner in which it originated; its funding sources and funding levels are two of the more important factors. Some museums depend almost entirely on public funds and give much importance to programs popular with the public. Others that depend on private endowments may be guided by the interests of their benefactors. University museums naturally emphasize collections and services of use to education. Few museums receive significant funds through charging for the services they provide, and so monetary profit is usually not a major factor in measuring the success of a museum - or of its information system.

The total number of objects held in museums nationwide is in the hundreds of millions. in 1973, the Association of Systematics Collections compiled a report [4] in which it estimated that there were somewhere between 200 and 300 million specimens of animals and plants in the nation's natural history collections, which are the largest of the four major types of museums. History museums probably house the second largest collections, art museums the third largest, and science and technology museums the fourth.

Collection sizes in individual museums range from small to vast. The American Museum of Natural History in New York City holds an estimated 30 to 40 million objects, and its neighbor across Central Park, the Metropolitan Museum of Art, houses probably three to four million objects. It is estimated that there are over 60 million specimens in the National Museum of Natural History in Washington, D.C. The massive size of collections poses the greatest problem in developing and maintaining information systems in museums.

Specimens come to natural history museums from a variety of sources for a variety of reasons. Private citizens donate specimens for tax purposes, to settle estates or to insure that valuable specimens are preserved. Zoos and botanical gardens donate deceased specimens. Also, some museums are designated by law as repositories for specimens collected by government agencies such as the U. S. Geological Survey. Seized contraband is often turned over to museums. Museums also purchase specimens. By far the most material is collected by scientists, however, working alone or on large expeditions or ocean cruises.

3. Data in Natural History Museums

3.1 Kinds of Data

The kinds of data recorded for specimens are determined largely by scientific requirements and collection management requirements. The following data elements are usually included, when applicable, in comprehensive records:

Collection management data

Museum name	Collection subdivision (birds, mammals, etc.
Catalog number	Donor's name
Number of items	Date of accession
Accession number	Date catalogued
Type of accession (gift, purchase, exchange, etc.)	Location in collection

Scientific data

Scientific name	Field number
Higher classification	Date collected
Author of scientific name	Sex
Year scientific name was published	Type of preservation
Bibliographic citation	Locality where collected
Type-specimen status	Locality number
Identifier's name	Altitude
Date identified	Depth
Collector's name	Stratigraphic position
	Ecology data

Many other data elements may be included which are specific to certain scientific disciplines. Examples are molt data for birds, chemical-analysis data for rocks and minerals, body length, hair color, and reproduction characteristics for mammals, flower color and tree size for plants, and cultural data for archeological materials. The number of observations recorded for specimens varies from just a few to hundreds depending on the conditions under which the specimens were collected and the amount of study they have undergone.

3.2 Role of Museum Records

Records about specimens are almost as valuable to museums as the specimens themselves. For the collection manager and for museum administrators, records are essential to the proper care and control of the collections and to the assessment of priorities for future collection growth and maintenance. To scholars, scientists and other users of the collections, records simplify the job of finding specimens, but of even more importance, records contain essential contributed data such as collecting dates and observations of other scientists which cannot be determined by examining the specimens themselves and which would be lost forever if not recorded. Natural history specimens for which no data are recorded are virtually worthless for scientific study and are often discarded or placed in expendable teaching collections, unless they are one of a kind, intrinsically valuable or unique in some other important way. Consequently, museums give high priority to maintaining effective records.

3.3 Conventional Records

The prime objective of a museum data system is to preserve the specimen data in a manner that it will not be lost and can always be matched with the specimen. The second objective is to provide whatever data are necessary to manage the collection and to keep the specimens safe, and the third objective is to provide indexes and other records which will facilitate the use and management of the collection. The data systems should also have these characteristics: (1) museum data do not become obsolete, so the records should be of archival quality in order to last an indefinite period of time; (2) years may pass between the points in time when a specimen is studied, but its records should be as readily accessible as those of specimens of current interest; (3) data about a specimen may be contributed at any time, so records must be updatable; and (4) original descriptions, geographic terms even if archaic or obsolete must be preserved rather than replace by modern expressions. Upgrading old data may change the original meaning.

The following five major kinds of documents make up a complete conventional record system in a natural history museum.

(1) Field documents: Some of the most important specimen data come to hand at the time the specimen is collected, eg. the date of collection, ecologic notes, and locality data. Extensive notes are recorded in field notebooks and on temporary (sometimes permanent) specimen labels. Several techniques are used to correlate the in the notebook with the specimens.

(2) Accession documents: Letters of transmittal, collection permits, shipping papers, and all other papers which come to hand when the specimens are transferred to the museum are kept. These papers are especially important in establishing that the specimens were legally obtained and are owned by the museum. Accession documents usually refer to groups of specimens rather than a single specimen.

(3) Permanent labels: Many curators regard the permanent label as the most essential document because it is always kept with the specimen and contains a complete set of the important data. Museums go to great lengths to assure that labels will last indefinitly and will not easily be separated from the specimens. As a backup measure, the specimen number is often written on the specimen itself to provide a link to the specimen catalog in case the label is lost or destroyed.

(4) Catalogues: If unique numbers are assigned to specimens, the data are usually recorded in numerical order in a bound ledger or on index cards, which constitute a catalog of the collection. The catalog often contains the most complete and up-to-date information about specimens.

(5) Indexes: Indexes to the collection usually consist of complete or abbreviated copies of specimen data on index cards arranged geographically, by object name, by donor's name, by culture etc. Good collection indexes can save considerable time in searching for specimens in a large collection. The internal systematic arrangement of the collection itself often provides an avenue of quick access also.

Most museums that use conventional documentation methods have not been able to maintain a complete set of records because they do no have the staff to handle the huge number of specimens. Some can barely find the time to put permanent labels on the specimens and to keep the collections in good order. The thought of cataloging specimens and of building indexes is out of the question. Accepted practice in botanical collections (herbaria) is to number the specimens serially, arrange them in the storage cases by their classification and let the collection itself serve as the catalog. Virtually the same is true in insect collections, but there the vast numbers of specimens (over 20 million in the NMNH alone) discourage even the assignment of serial numbers to the specimens. In some collections where specimens typically are less numerous, eg. mammals, birds, and reptiles, catalogs may be kept for all specimens or at least for the most important ones, but indexes are less common.

Most scientists who regularly visit various collections to conduct their research expect to do extra work to overcome incomplete documentation. A scientist is often faced with having to mount, as it were, a new expedition into a collection to find the desired data. The potential value of many significant collections is not being realized because they are so inaccessible owing to minimal documentation.

3.4 Data Processing Systems in Museums

A few museums began looking into data processing about fifteen years ago to solve their data-handling problems, and several software packages designed especially for museum use have been put together by museums, universities, and more recently by commercial firms. Data processing was used at first by only the few museums that had the right combination of hardware, expertise, and management to overcome the considerable start-up and growing problems. But in recent years, improvements in technology, decreasing costs, better understanding of the role of data processing and other positive developments have induced a general surge of interest across the entire museum community. It appears that data processing will be commonplace in museums in just a few years.

Most museums are still at the stage of comprehending or assimilating the basics of data processing. The questions commonly asked now by museums are: "Should we install a system," "What systems are available," "Which system should we use," "Can we afford to use data processing," and "Where can we get help in getting started?" Many grant proposals have been submitted by museums to funding agencies in recent years seeking support for installing systems, and some of them have been accepted. For some of the less fortunate museums, however, data processing is out of the question, but many of them are tuning up their conventional record systems to allow easier conversion to automated systems later.

The National Museum of Natural History (NMNH) was one of the first museums to begin using data processing. It's system will be briefly described here to give the reader an idea of what a museum does with a data processing system.

NMNH began using data processing in the early 1960's with two basic objectives in mind: (1) to reduce the amount of work required to catalog and label new specimens and (2) to build an machine-readable data base that would allow extensive cross-indexing of the existing collections and rapid retrieval of combinations of specimen-associated information. With the help of the Smithsonian computer center, the museum designed and wrote it's own batch-processing software, which has since been put into use at several other museums. A central office was set up to coordinate the data processing program, but it was left up to the scientific departments to determine which of the millions of specimens to process first and which data to enter for them.

Because it is unlikely that NMNH will have enough time and money to process all of the data for the entire collection of 60 million objects, priorities for processing are crucial. Some departments have given priority to using the museum's system to catalog new specimens. An average of twenty to thirty data items are entered for each specimen from a variety of source documents such as field notes, temporary labels, and data sheets. Several techniques are used to eliminate redundant typing, and automatic typewriters, the computer, or programmable terminals are used to print multiple copies of specimen labels, index cards and other specimen documents.

Other departments have stressed retrospective processing of data on important older collections such as type specimens, other intrinsically valuable specimens, or specimens of current research interest. Conventional catalogues or indexes that already exist are sometimes used as the source documents, their internal systematic arrangement serving to reduce redundant keystroking. In some cases, to spread the data processing resources over a larger portion of the collection, only the main data elements are recorded rather than all of the elements typically found in a catalog. The data base then serves as an index to the collection rather than a catalog of the specimens.

NMNH has entered data for over 3 million specimens so far and is adding data at the rate of a quarter of a million specimens per year. Costs for processing the data are similar to the costs of conventional systems, but the benefits in terms of indexing and retrieval are far greater. Here are a few of the products already provided to users of the collections:

* Computer-plotted map of squid-collecting sites in the southern hemisphere.

* Culture/locality/object-name index to an ethnology collection of a quarter of a million objects.

* Computer-produced microfiche (master list, species index and locality index) for records of ticks and their hosts. This three-quarter-inch stack of microfiche enabled a scientist to take the data from 80,000 index cards on an extended visit to the British Museum to restudy some large collections.

* List of sediment samples collected along stretches of the Potomac River specified by latitude/longitude coordinates. The list was provided to an environmental consultant.

* Published lists of type specimens and other important specimens held by the museum.

An here are a few of the products provided for collection management purposes:

* List of ethnology specimens accessioned as gifts.

* List of specimens missing from a type collection (probably misfiled with non-type specimens).

* List of specimens with duplicate catalog numbers.

* List of mineral specimens on exhibit.

* List of specimens lacking mandetory items of data.

* List of Seminole Indian objects; produced as a turn-around document to be checked against the collection.

Some of these products could be provided only after searching tens of thousands of records, a feat which would have been impossible in a conventional data systems.

Data processing systems in other natural history museums are similar to the one at NMNH in their basic objectives and data sets but differ in details. For instance, the Carnegie Museum of Natural History in Pittsburgh is using commercial software and does its processing on a vendor's system. At the University of Michigan, herpetology collections are processed on a university system which is at least partially interactive, and the same is true for archeological collections being catalogued at Arizona State Museum. One of the most promising systems is in Canada, where the government is financing a network for the national museums of Canada. The data processing center is in Ottawa, and is accessible to the participating museums via terminals. A couple of museums have experimented with minicomputers.

3.5 Data Problems

The characteristics of museum data and the uses to which they will be put pose some interesting challenges for museums in defining their data standards and in operating their data processing systems. The following four examples explain some of the more severe problems.

a. Problem of Old Records

One of the authors (Shetler) recalls that when he came to the Smithsonian in the early 1960's, he was handed a small packet of papers with a tag attached on which was written "Hold for three months and then throw out." The date on the tag was 1926! This recollection illustrates a major problem that faces museums on a daily basis - the condition of the older records.

Conventional museum records which have been accumulated over periods of 50 or 100 years contain just about any type of data problem imaginable - missing data, misspellings, transcription errors, poor handwriting, faded entries, obsolete terminology, synonyms, unexplained codes and marks, vague or general terms, transposition errors, obsolete geographic terms, inconsistent formats and so forth. The wise user of older museum records makes no assumption about their accuracy unless some kind of verification is possible.

Obsolete or vague geographic terms are of particular concern in museum data processing systems. Though scientists today require precise locality data for specimens, many older specimens are simply labelled "Dakota Territory," "Colorado River," "Great Britain," or "Phillipine Islands," and even newer specimens may have been collected in countries that no longer exist. Museum data standards must take into account such historical aspects of data. Most museum curators require that old locality data be entered exactly as given, since attempts at clarification - if possible at all - may change the original meaning of the data. Despite rigid controls, however, errors do creep in, as in one case where "Oklahoma Territory" was changed to "Oklahoma," cutting several thousand square miles off the original locality designation.

The eventual outcome of processing older data in museums is that the data are either entered "dirty" or a great deal of time is spent in weeding out the problems. Which direction a collection takes depends on its priorities. For instance, if a museum department is preparing to publish a catalog of the important specimens in its collection, accuracy is of paramount importance. On the other hand, if auditors have ordered that a computerized inventory of a collection be prepared in a short period of time (this has happened), less accuracy or highly abbreviated data will be tolerated, and the problems will be cleaned out later.

b. Problems with New Data

During the long history of collecting, scientists have come to follow fairly standard practices in the kinds of data they record and the kinds of original documents they prepare, but the standards do not always provide the precise definitions of terms, syntax, and formats of data required in a data processing system. Consequently, original data coming to a museum with new specimens may be presented in a wide variety of forms, which complicates the museums's job of curating and documenting the specimens.

Data processing has not been in museums long enough for its strict data requirements to become known and followed by all collectors, but progress is being made on many fronts. For instance, many scientists (eg. Erwin [3]) who are using data processing in their research have developed standard forms

for the collection of data in the field, and follow their standards consistently. When the benefits those scientists have derived from standardization become well known, other collectors may adopt their procedures.

c. Classification Problems

One of the oldest and most formal data standards followed by museums is the binomial system of nomenclature which dates back to the eighteenth century and which requires that the scientific names of animals and plants be made of two words, a generic name and a specific name. Assigning the correct scientific name to a specimen is just one part of the process of classifying organisms, however. Scientists are also concerned with higher levels of classification where genera are grouped into families, families into orders, orders into classes and so forth. The entire classification system is meant to express the natural order that exists in nature.

In the ideal data processing system, the correct scientific name for a specimen would be entered and would never need to be changed. A higher classification system agreed upon by all scientists would also be entered to facilitate retrieval of larger related groups of specimens. But the ideal system is virtually impossible to achieve. The classification of animals and plants is a highly subjective process which is in a constant state of flux. New scientific names are constantly being created, and existing names are being changed or replaced as scientists discover new facts that change their interpretations. Higher classification schemes are almost as numerous as the scientists using them, and scientists often have difficulty in agreeing on the proper identification for specimens. Museum data standards must have enough flexibility to take these fluctuations into account but also enough rigidity and precision to enable classification to be a useful point of retrieval for museum data banks.

d. Problem of additions to the minimum data set

Data processing opens up many new ways for using museum data, but these may in turn require museums and collectors to add to the list of data elements they normally record. Computer mapping gives a prime example.

A major line of investigation of animals and plants is to study their distributions to determine such things as what environmental factors control the distributions and how the distributions have changed in time. The advent of computer mapping promises to greatly speed up distribution studies. Museum collections represent the largest sets of distribution data available, and are particularly important because they include the time dimension which allows scientists to determine distributions in times past. The scientists' interest in using specimen data for such studies poses problems for museums, however, because machine-mapable coordinates (eg. latitutde and longitude) are available for only a minority of the specimens. It is unlikely that museums ever will take the time to plot coordinates for their older specimens, except perhaps by using the computer to match their data bases with geographic data bases (obsolete and imprecise geographic terms would be problems in such a process), but at an increasing frequency, museums are taking the extra time necessary to record mapable coordinates for new specimens.

4. Coordination of Data Standards between Museums

The scientific disciplines such as entomology, archeology, botany, paleontology and mineralogy have a vital role in setting the course of data processing and data standardization in natural history museums. Their members are by far the major users and contributors to the collections, and most museum curators and administrators are scientists themselves. The professional societies such as the Society of Vertebrate Paleontology and the American Society of Mammalogists are often the major forums for inter-museum discussions concerning such things as proper curation techniques, national priorities for collecting and standards. In recent years, many of the disciplines have formed advisory committees, at least partially funded by the National Science Foundation, which have prepared white papers outlining the needs of their collections across the nation. The need for improved documentation systems has often been identified.

Some of the disciplines have given much consideration to data processing. The American Society of Mammalogists has probably gone the farthest in establishing a program called the National Information Retrieval Network for Mammalogy (NIRM). A goal of NIRM is to build a data base for the nation's mammals collections, which are estimated to hold a total of 2.5 million specimens. Among other things, the combined data base would help a mammalogist discover where specimens of interest to him/her are located, and it might be used by museums to plan exchanges to bring all specimens of one kind or from one geographic area together.

NIRM has identified a minimum data set and has written standards for the data elements. Though funds have not been obtained to operate NIRM on a nationwide basis, many museums which are using data processing in the mammal collections are following the NIRM standards, and their progress reports are included in the NIRM report periodically prepared by ASM.

Museum organizations such as the Association of Systematics Collections, the American Association of Museums, and the International Commission on Museums are also taking steps that may assist coordination of data standards between museums.

Although most museum data processing standards have been written to meet the needs of individual collections, there is great similarity between standards of different museums because of the inherent similarity between collections, the strong coordinative effects of the scientific disciplines and museum organizations, the sharing of software, data standards and data processing experiences between some museums, and the strong desire in many museums to use systems compatible with others. It seems likely, therefore, that many museums will be able to combine their data bases to produce composite indexes to their collections and other useful products that would facilitate collection planning, maintenance and use. But, this assumption has not been tested, since examples of such inter-museum data exchanges are almost non-existent. Museums need to act soon, before their data bases are too large, to carry out at least experimental exchanges to discover where modifications to achiever greater compatibility are needed in their data standards.

Whatever way coordination of data standards is achieved between museums, experience suggests that the process should have the following characteristics to be most successful:

(1) The standards must serve a current practical need; standards which are intended to anticipate future needs will probably be used less.

(2) The standards should not compromise scientific freedom of thought.

(3) The standards cannot require an expensive retrospective change in existing data banks and other collection documentation.

(4) The standards should bring benefits soon to individual museums.

(5) The standards should be easy to adopt or highly beneficial if they are complex.

(6) The standards cannot require elimination of essential information, i.e. reduction of the minimum data set.

Above all, more communication and sharing of ideas, and a greater awareness of the potential of data processing in museums will be needed.

5. References

[1] American Association of Museums, The Official Museum Directory, National Register Publishing Co., New York, 891 p. (1977).

[2] Chenhall, Rober G., Museum Cataloging in the Computer Age, American Association for State and Local History, Nashville, 261 p. (1975).

[3] Erwin, La Verne J. Magarian, Application of a Computerized General Purpose Information Management System (SELGEM) to a Natural History Research Data Bank (Coleoptera: Carabidae), The Coleopterists Bulletin, 30:1-32 (1976).

[4] Irwin, H.S., Payne, W.W., Bates, D.M., and Humphrey, P.S. (Compilers and editors), America's Systematics Collections: A National Plan, Association of Systematics Collections, Lawrence, Kansas, 63 p. (1973).

[5] Shetler, Stanwyn G., Demythologizing Biological Data Banking, Taxon, 23:71-100 (1974).

[6] Squires, D.F., Data Processing and Museum Collections: A Problem for the Present: Curator, 9:216-227 (1966).

[7] Vance, David, Computers in the Museum, IBM Corporation, 69p. (Available through IBM representatives) (1973).

Keynote Address

Stephen M. Gershenson
Assistant Director
Commission on Federal Paperwork

Introduction

Today's headlines scream "hot war," "cold war," "crime" and "scandal." They blast out that this is a world beset by personal conflicts and profound, disquieting political and social turmoil. My remarks this morning have nothing whatever to do with these cosmic concerns. Rather, they are modest in scope, tempered in rhetoric, and confined to an examination of four questions:

- What was the Commission on Federal Paperwork instructed to do under the law that created it?
- How did the Commission go about its work?
- What did the Commission find? What are its central conclusions?
- What, in brief compass, are the Commission's recommendations, its agenda for reform?

The answers to these questions form the nucleus of the Commission's Final Report. This report, transmitted today to the President, the Vice President and the Speaker of the House of Representatives, is both epilogue and prologue. It is an epilogue for the Commission on Federal Paperwork, its mandate fulfilled and its tenure expired. It is prologue for the work yet to be done.

This Symposium is, for several reasons, an appropriate forum in which to share the Commission's work. First, the Commission's origin and early thinking owes much to the conceptual and technical contributions of your speakers and attendees. Second, the Commission conferred with several of you to sharpen and refine its analyses and investigations. Third, and most important, many of you will be asked to implement specific proposals made by the Commission. Entrusted with this responsibility, you ought to know the source and evolution of those recommendations.

The Commission's Mandate

The first question, "What was the Commission instructed to do?" has a three-part answer. One important objective of the Commission, from my viewpoint, was to surmount its name. The law creating the Commission[1] uses the word "paperwork" only three or four times. There are, however, repeated references to information, e.g., "Federal information reporting requirements," "information reporting burden," "information policies and practices, "...management and control in information activities," "Federal information needs," and so on. In name, we were the Commission on Federal Paperwork; more accurately,

[1] Public Law 93-556; 88 Stat. 1789.

we functioned as a commission on information management. This was done deliberately. The Commission was instructed to "...study and investigate statutes, policies, rules, regulations, procedures, and practices of the Federal Government relating to information gathering, processing, and disseminating, and the management and control of these information activities."[2]

More specifically, the Commission was to consider:[3]

- the nature and extent of current Federal requirements for information from other public and private entities;
- the effect of existing statutes on the information requirements of the Federal Government and authorities of existing Federal agencies to collect information;
- the nature and extent of management and control over the determination of Federal information needs and the choice of information gathering, processing, and dissemination methods;
- the nature and extent to which Federal agencies cooperate with State and local governments and private agencies in collecting, processing and disseminating information;
- the procedures used and the extent to which considerations of economy and efficiency impact upon Federal information activities, particularly as these matters relate to costs burdening the Federal Government and providers of information; and
- the ways in which policies and practices relating to the maintenance of confidentiality of information impact upon Federal information activities.

The objectives of the Commission were not only to study, but also to reform. Our charter required us to identify and articulate those changes, "possible and desirable," in Federal information activities which would achieve several specific goals:[4] First, to assure that the informational needs of Federal officials, or those acting in their must furnish information to the Federal Government; third, to develop "appropriate standards of confidentiality" for information held by private citizens or the Government, and the release thereof; fourth, to provide that information held by the Government is processed and disseminated to maximize its usefulness to all Federal agencies and the public; fifth, to reduce the duplication of information collected by the Federal Government and by State and local governments and other collectors of information; and, finally, to reduce the costs of Federal paperwork.

The requirement to investigate Federal information management policies and practices, coupled with the obligation to develop recommendations, gave this Commission the broadest mandate since the Hoover Commission's to examine the programs, practices and processes of the Federal Government. The key elements of public management were to be refracted through a paperwork prism.

---

[2] Ibid., Sec. 3. (a)

[3] Ibid.

[4] Ibid., Sec. 3. (b)

The Work of the Commission

Given this charter, and a lifespan of two years, how did we go about our work? We went about it in four complementary ways.

Commission Hearings and Advocacy. The Commission held a series of hearings throughout the nation -- 25 days of hearings in 19 cities. So far as we know, no other temporary commission has been so diligent in seeking out citizen and business concerns, nor so attentive to the issues raised. Often, these hearings led to what we refer to as "weed cutting" exercises in which individuals who testified before the Commission presented "horror stories" confined usually to a single form or report, garbled instructions, conflicting regulations, duplicate requests for information and so on. To respond promptly to these complaints the Commission created an advocacy/ombudsmen unit. We had a toll-free "hot line" open 12 hours a day. Complaints were received, documented, and investigated; alternative solutions were developed in consultation with the agency or agencies involved. The preferred option was then communicated to the agency through recommendations formally adopted by the Commission.

Impact Studies. Identification of problems from the respondents' point of view occurred through hearings, correspondence, and contacts with representative organizations. The Commission's impact studies synthesized this information and undertook additional research to report on the burdens of paperwork and red tape from five types of respondents: farmers, individuals, labor organizations, large and small businesses, and State and local governments.

Program Studies. Eighteen program studies focused on the paperwork and information management problems arising from either a single piece of legislation -- such as the Occupational Safety and Health Act or pension reform (ERISA) -- or a broad Government activity, such as housing, energy and welfare. Individually, these studies yielded immediate improvements; collectively, they provided substantial evidence which illuminated the systemic problems of Federal paperwork.

Government-wide Studies. The fundamental institutional reforms sought by the Commission were explored through thirteen process studies. These studies analyzed problems that cut across organization boundaries, were not limited by specific program legislation, and had persisted for some time. In purpose, these studies sought changes in those statutes and policies which serve as the bedrock of Government management procedures affecting paperwork requirements. Topics of this kind include treating information as a resource; central clearance of reports and information requests imposed on respondents; legislative barriers and administrative constraints on information sharing; information exchange between and among levels of government; and the role of Congress in creating and controlling paperwork.

Emerging from these studies is a body of evidence that is at once broad in scope, rich in detail and resistant to easy summarization. Nevertheless, there is an obligation to summarize and simplify and it is to that task that I now turn.

Findings and Conclusions

The principal findings and conclusions of the Commission can be sketched quickly. The first cluster of findings are general in nature; the second focuses on the key causes of excessive paperwork and red tape.

General Observations. I will confine myself here to only a few general observations, mindful of Justice Holmes' admonition that an examination of the obvious is often more valuable than research into the obscure. What now is obvious is that:

- Paperwork originates in the political, not the administrative process, i.e., legislation is the root cause of paperwork. Surveys, applications, reporting and recordkeeping requirements are either mandated or authorized by law.

- The scope and diversity of Federal Government functions are mirrored in its information-gathering activities. As the number and type of Federal programs grow, so too does the demand for more kinds of information at increasing levels of detail. Federal agencies, and their State and local counterparts, now collect information from individuals, commercial and non-profit enterprises, educational institutions, labor organizations and from one another. This information is used for such functions as licensing, registration, insuring, training, regulating, servicing, diagnosing treating, charging, paying, or conveying other benefits or penalties. Paperwork is the inevitable if unwelcome consequence of the demands we ask Government to satisfy.

- Most of the paperwork required by Federal agencies is necessary and important to the planning and operation of authorized programs and functions. More pointedly, I would argue that "routine paperwork" is often the most efficient, most economical and least intrusive way for the Government to get the information it needs. It is by far preferable to permanent, on-site inspectors or auditors, special or one-time surveys, litigation and its discovery procedures, and legions of enumerators.

- The ways in which the Federal Government manages (or mismanages) its information resources can be improved. The opportunities are abundant, visible and recurring. The mismanagement of information is not a situation to be endured, it is a problem to be solved.

Key Causes of Excessive Paperwork. The Commission identified seven basic causes of, or factors which contribute to excessive paperwork. The first cause, according to our Final Report, is poor communication. Poor communication in the sense that government is obscure to its citizens -- impenetrable, remote and adversarial in nature. This finding, although easy to state, has important consequences.

> Traditional democratic theory presupposed an immediate and evident relation between the individual citizen and the government.... [T] he state was to confront the citizen directly as both servant and master. The issues debated in the legislature [and, by extension, the decisions of executive and regulatory agencies] would be comprehensible to every educated subject and their relevance to his interests easily understood.[5] (Bracketed material not in the original. S.G.)

These presuppositions no longer hold; in fact, they are historical curios in an age which needs permanent, complex institutional arrangements to transmit the "will of the people" to the elected governors, to refine, adjust and clarify the mix of rights and obligations. The strategy for improving communications between citizen and government is, therefore, not laid out for us in bold letters. Rather, it is marked by careful attention

[5]Robert Paul Wolff, et al. A Critique of Pure Tolerance (Beacon Press, Boston, 1965) pp. 6, 7.

to details, sensitivity to subtle distinctions and nuance, and tedious work.

A second cause of excessive paperwork is the insensitivity of public officials. We include here both elected and appointed persons who, in preparing their informational requests from the public, are heedless, if not necessarily mindless, of the burdens that are imposed upon the public. The cost of complying with informational requests in inadequately, and in some cases never considered.

A third cause of citizen concern is the incomprehensible nature of some of our forms and instructions. When college-level reading skill is needed to decode or decipher the simpliest IRS form, something is amiss in Washington.

More important, and frequently encountered, we found overlapping organizational structures and jurisdictions to be a principal spur to Federal paperwork. For example, the Farmers Home Administration, the Veterans' Administration, and Department of Housing and Urban Development guarantee or insure loans for private homes. The information needed by the lenders for each of these programs is more or less identical. Nevertheless, there are abundant and redundant forms that collect the same information, that double, triple and quadruple the burdens upon the lenders and those who seek loans. Here is another example: prior to the new Department of Energy, now in embryonic form, at least six major executive agencies collected information on oil imports. At the end of each month or quarter, when the data were displayed, one simply could not reconcile the findings of one agency with those of another. In some cases definitional disputes arose -- that is, the same term or phrase had more than one meaning. In others, reporting dates differed from one system to another. When a tanker enters a harbor, the Bureau of Customs receives a manifest that lists the volume of imported oil as of that day. However, the refiner who has not yet taken possession will report a different date for the importation of the oil. In result, we have systems that are inconsistent and that produce contradictory information on the same subject for the same time period. No wonder that citizens are baffled, and planning and policy decisions cannot be made with anything approaching a high order of confidence.

A fifth leading contributor to paperwork is inadequate program design. Federal agencies responsible for designing, operating and evaluating a program sometimes strain belief to produce inefficient, ineffective, cumbersome, costly, and intrusive reporting systems. These are exercises in imagination rather than in self-discipline or common sense.

Closer to the concerns of this Symposium is the finding that poor information management practices prevail. The Commission is mindful of the slow development of information standards, the difficulties in information transfer, the absence of an inventory or directory of Federal information holdings, the problems in the clearance process, etc. We have documented several instances in which reporting systems were "enriched" or enlarged not because of need but because computer time was available. We have corroborated the observation that computerized applications expand to meet the capacity of the configuration. Stated bluntly, technology has outrun wisdom.

A seventh cause of paperwork problems can be traced to inconsistent, incomplete and ineffective laws, policies and practices dealing with confidentiality and privacy of information. The two statutes that control here -- the Freedom of Information Act and the Privacy Act -- are each in their own way, limited. The Freedom of Information Act enunciates a policy of broad public disclosure for all types of information, and then proceeds to list nine classes of (discretionary) exemptions. The Privacy Act, on the other hand, deals with only one category of information -- personal/individual data -- but treats it through collection, storage, processing, dissemination and use.

The confusion that results contributes to and reenforces compliance machinery that is inadequate in scope, diffused in purpose and meager in results.

Overview of Commission Recommendations

So far, we have answered three of the four questions posed at the outset. We have outlined the Commission's mandate, its work program and its findings. The remaining question is, "What does the Commission propose? Where do we go from here?"

I think there are about a half dozen or so categories of reform that we should seek. The first two I would call "dispositional," in the sense that they require conceptual and attitudinal changes. The conceptual change is one that you have already made. It is testified to by your presence here -- part of a small but growing group that recognizes that information is resource, not a free good. Not only is information a resource, it is a manageable resource -- one that can be planned, one that can be organized, one that can be controlled, one that can be evaluated, one that can be budgeted. In essence, information can be treated as a resource in much the same way we now treat personnel, real property, space, equipment and supplies. This is not a radical notion, but it is one that is resisted by many people in government because they have never been asked to account for the information practices they have committed.

Along with this conceptual change -- that of treating information as a resource -- there is an attitudinal change that we seek. Historically, administrative and management reforms offered by study commissions have emphasized economy and efficiency in terms of the internal machinery of government. Such proposals reflect implicitly certain beliefs about the role of government in a laissez faire environment. The fact of the matter is that the Federal Government is now enmeshed in the day-to-day lives of our citizens, our institutions, and our organizations in unprecendented ways. The National Government has shifted from a passive referee or arbiter of private sector disputes to a provider and deliverer of services. We argue that public management or public administration needs a new body of doctrine more suitable to today's realities than to yesterday's rhetoric. We call this concept "service management."

Service management is the set of principles, procedures and techniques necessary if Federal programs involving those outside government are to operate according to these long-standing tenets of public policy:

- Government exists to serve and protect;
- Federal programs should be conducted with economy, efficiency and speed, coupled with fairness and equity.

Our proposals are consistent with, and build upon existing management techniques by adding two distinctive disciplines:

- Managing information as a resource; and
- Analyzing alternative ways of organizing and operating programs so that costs and advantages are appropriately distributed among all parties involved.

The concept of service management must be understood as preparing government to manage not only its costs for Federal programs, but to manage as well the costs which Federal programs impose on others.

I had mentioned earlier and gave some examples of organizational anomalies as causes of paperwork. The Commission therefore finds it necessary to recommend a number of organization changes which involve both the Congress and the executive agencies. For the Congress the Commission has recommended that the House Committee on Government Operations and the Senate Committee on Government Affairs assume the leadership role for all legislative oversight for the information policies and practices of the Federal Government.

Within the executive branch, the Commission has found the existing machinery of Government for coordination, policy and standard-setting, and enforcement to be inadequate and occasionally embarrassing. The Commission has recommended that the President and the Congress give prompt consideration to establishing a central, Cabinet-level Department of Administration to permit improved policy formulation, policy oversight, and standard setting in addition to specific operating and coordinative responsibilities.

There are also many other recommendations made by the Commission. They are less sweeping in nature and certainly less glamorous. We have made some specific proposals to change certain procedural, operational and technical aspects of government. We have, for instance, urged changes to the reports clearance process -- the system whereby agencies now submit (to either OMB or the General Accounting Office) requests for information from the public. The clearance process is marked by split jurisdiction between the OMB and the GAO, late arrival on the scene, a slow pace, and an adversarial nature. We have recommended ending this split jurisdiction; we have suggested changes in the way Federal agencies manage their records - physical records and reports. There are also some changes that the Commission urges in terms of technical consideration; they involve data standards, data element names, definitions, codes, improvements in the way in which Federal agencies design forms, readability standards, compatibility, information sharing, data transfer techniques and technology.

The last type of change is what I would call an "economic change." If we understand information as a resource, if we understand it as a manageable resource, there ought to be some way in which we can account for it. The Commission has produced some suggestions, some thoughts, perhaps even a little bit of wisdom, on an "information resources management" budget object classification code.

Conclusion

The temptation is great for commissions to serve as kind of hit and run drivers. A commission, at the end of its tenure, points with pride at its own accomplishment and views with alarm the work undone. There is a paradox of timing in commission reports. Commissions are most visible when they deliver their final report -- the very instant their authority, capacity, and clout to get anything done vanishes. We are more fortunate because the statute that created us requires the Office of Management and Budget to submit, at six month intervals for the next two years, reports to the Congress on the status of Commission recommendations, and to propose needed legislation to implement them.

I think it is appropriate to recognize that poets often have more insight than analysts. I outlined for you the bold and broad mission of the Commission, and was reminded of one poet who said, "A man's reach should exceed his grasp, else, what's heaven for?" I suspect that we have not necessarily achieved all that the Congress or the President desired of us. On the other hand, we can take some comfort and some counsel from the words that Ulysses, in his old age, gave to his colleagues -- men who had been with him since the Trojan war. He observed that they were no longer young, no longer vital, no longer vigorous, but still it should be their resolve "to strive, to seek, to find and not to yield." It should also be ours.

Standards for Electronic Data Interchange

Edward A. Guilbert
President
Transportation Data Coordinating Committee

"Standards - standards - everywhere, but not one to fit my needs." That seems to be a hue and cry that is heard in the marketplace of industry and in the clinical corridors of the government agencies. How often have you heard the phrase "What we need is a standard - any standard - so that at least we have a common specification to work from." Then, the first meeting is called and those companies that have the budget latitude enabling them to send one representative for one day gather at 10:00 a.m., break for lunch at noon, reassemble at 2:00 p.m., and break at 4:00 p.m.

Out of the luncheon portion of that meeting comes the standard recommended by the six participants. Out it goes to hundreds of others who are often too busy with other things to send in their vote. And, if they do send in a "no" vote, they offer no alternatives.

This is certainly not the way the National Bureau of Standards, ANSI, nor ISO function. But we did find it to be a trend in some of our industry activities. It was to overcome that type of standards development and coordination that the founders of the Transportation Data Coordinating Committee banded together and established a full-time staff capability to pursue standards for the electronic interchange of data in the transportation/distribution communities. I am part of that full-time staff and that is why I am here today to tell you about the Transportation Data Coordinating Committee, its goals, programs, activities, and accomplishments. Of course, I shall do this in a very brief, modest, and humble manner.

The TDCC is not a household word. As a matter of fact, it is a very select, low profile, "get the job done right" organization. It is a non-profit center supported by manufacturers, shippers, railroads, motor carriers, ocean carriers, airlines, forwarders, banks, and computer and communications activities that are dead serious about cutting administrative costs, enhancing productivity, and making maximum use of current and future communications and computer technology.

Right now, each of you are asking yourself "I wonder if my company belongs to this dynamic organization because we are interested in the same objectives?" Make a note to find out when you report back to your managements on this conference. Whether your company belongs or not, plan to keep informed of our programs because we are on the glide path to new technological innovations that will impact on the way your company transacts business in the future.

Now that you know who we are, let me tell you what we are and what we are doing and why standards are so vitally important. First, we must define the problem area that we are dealing with. For example, our members are operationally oriented rather than theoretically or academically or concept oriented. Why do they need standards? Let's cite an example. From the order placement cycle, data is entered into their computers. This accommodates the internal company program requirements for production, inventory, scheduling, and other activities. But, when the goods are to leave a plant and go to the external environment it becomes necessary to turn the computer into a printing press to crank out thousands of bills of lading needed to turn the shipments over to the carriers. The carriers take these reams of

documents to their offices where a battery of coding clerks must code the shipment details and keypunch the data into their computers. They do this costly and time-consuming task to create the data base for their operation and to rate the shipments. This process then creates thousands of freight bills produced in three-to-ten copies, which must be enveloped, addressed, and mailed back to the shipper. The shipper receives, opens, and codes all of these bills, enters the appropriate data back into their computer to validate shipments, check the rates and charges, and activate an accounts payable authorization. This starts the next paper cycle which is the check writing. Thousands of checks are mailed to the carriers or their payment agent which must be checked against their accounts receivable. Checks are deposited in their bank, cleared, and routed to the shipper's bank and ultimately mailed as cancelled checks back to the shipper. Now I ask you -- in an age of sophisticated, low cost data systems technology -- is it practical for a shipper to resort to manual paperwork processing to accomplish a data flow between computerized systems? Even more parochial is the fact that the data that the shipper gives a carrier on bills of lading is the exact same data that the carrier gives back to the shipper on freight bills. This is but one example of the institutional idiosyncrasy that the paperwork syndrome has burdened the business community with.

Let me cite just one more example so that you will totally grasp the importance of standards. These are actual company experiences that I will describe. Company A places 5,000 orders for products per month with Company B. The orders are computer-generated but must be printed out and mailed to Company B. Company B must open envelopes, code and keystroke order data into their computer to determine whether it is a production or inventory item. When the shipment is made, an invoice must be prepared on each shipment. These are mailed to Company A (again it is the same basic data that Company A gave to Company B). Company A opens envelopes, codes, and keystrokes the invoice data to validate the order requirements and prices. They activate accounts payable instructions and the check writing process begins. They mail checks to Company B or his lockbox and again the merry-go-round of the check passing cycle begins, ending with the cancelled checks back with Company A.

Here is the message! With agreed upon standards, business systems transactions can be accommodated more effectively, more efficiently, more economically, and without human error by employing Electronic Data Interchange directly from one party to another. No privacy, security nor data center problems since we are simply changing from manual to electronic transmission.

The "Preliminary Functional Specification for a Prototype Electronic Data Interchange System" was produced by the TDCC as part of a contract with the Office of Facilitation of the Department of Transportation. It creates the means of electronically exchanging shipping, billing, tracing, audit, payment, and export and import data. It contains the data elements and definitions, message formats, transaction sets, segments and segment identifiers for both domestic and international trade and transportation. The document reflects the concentrated efforts of some 180 industry and government experts that reached agreements which enabled this specification to be used for pilot program implementation, not on a test but on an operational basis. In addition, TDCC has produced a General Communications Specification, General Programming Specification, and General Systems Specification.

TDCC is not in a "Users Group" posture, with Electronic Data Interchange programs for several different applications not being planned, programmed or already implemented. Obviously, until we have live data experience - subject to full impact analysis - we choose to remain in a pilot program status for the next two years. We still have much to perfect, particularly in terms of code improvement, interchange agreements, and tariff modernization.

About this time you are probably thinking: "Wait a minute, Guilbert. You've made a pretty fast trip with an interesting scenario. You've given us very little on some very sticky technical problems. How about some technical backup on the approach?" First, let me give you some points of reference which relate to various technical problems. In the end, you should conclude that we have no major technical problems. Nor do we foresee major security or legal problems. If we have any problems, it is in the area of company priorities in terms of resource allocation which is necessary for progress.

At the outset, you must recognize that each company selling or buying goods or services forms an environment within which control can be exercised. Let us call that the internal environment. In the case of systems, it is an internal systems environment. The style of management dictates the character of the environment. Some companies completely delegate the management responsibility to the division, plant or group level in a kind of stand-alone cost center. Each has its own resources and directed profit objectives. Other companies delegate production and centralize financial management. Still others operate under highly centralized concepts of management. There are many approaches and all have proved to be successful. It depends on the type of business and the personality of the management enclave.

The point is, the management concept more often than not dictates the character of the internal systems environment -- some formed by "islands" of completely different systems, and others with highly centralized systems. So we cannot approach this thing we call electronic data interchange with any all-inclusive set of requirements or a single systems concept. We must keep the approach loose, with a basic set of principles and some generalized standards. We must depend on evolution to change business practices -- not revolution. We must parallel proven practices in the application of technology -- not cause an upheaval. We must exploit technology to benefit the business -- not to prove a concept. We must allow the market for electronic data interchange to develop along the path of least resistance -- not force it into a mold with exotic schemes.

So much for philosophy. How about some principles to back up the TDCC theory of EDI in transportation and trade?

Initially, we must conclude that goods are manufactured, transported, and delivered - and money is exchanged as a result of, and on the basis of, information flowing from one party to another. Information is our target. Information is a resource and, in the business world, it is recycled over and over - sometimes in the same form - sometimes in modified form - and sometimes in consolidated form.

Where does it start - where does it originate in reference to the physical movement and delivery of goods? This leads me to our first principle. Nothing happens in business without some form of external stimuli. A company must get an order before any positive action can be taken to move goods. A transportation company cannot perform a single act of transporting without a service order or instructions. The same applies to forwarders, consolidators, brokers, and clearinghouses. Further, the money repository - the bank - takes no action regarding the movement of money without instructions. As you can see, therefore, this necessary outside stimuli is the principle of information flow that led to the current condition of a paper-clogged world. Paper was the best and only media of communications available before the advent of electronic digital systems technology. Paper served as the means of crossing the frontier between respective systems environments of the originator and the recipient. Paper is still the principal means of communicating between environments.

Because each system is controlled by the company, each has assumed a different character with inherent or manufactured standards in terms of methods and procedures. That process of systems development has carried over and is reflected in the electronic systems of each company. Comparatively speaking, we have evolved a multitude of equal but unlike systems. This is where some problems begin to surface. How do you move from the master/slave relationships common within internal systems environments which are under the complete control of a single entity to a broader environment of master/master relationships between equals? We did it by concentrating on data and its structure, software, and communications. We established standard methods, procedures, and structures. The TDCC standards are highly generalized which provides great flexibility for modification, expansion, and responding to a very wide spectrum of requirements. At the same time, rigid standards were set for data identification, edit, and control. A standard language was created for communication between unlike systems environments while, at the same time, facilitating the translation from the interchange language to internal data structures. TDCC focused only on the interface problem since the internal systems characteristics are unique to each party. We produced a General Programming Specification which expresses the data standards in programmer terms. It offers a table-driven approach that does several things. It allows us to easily modify the standards -- such as introducing new applications, new data elements -- without having to change the programming system. The system provides an efficient way to assign incoming data to internal system locations and/or translate expressions or codes from the external standard to those used internally. TDCC fully recognizes the importance of a standards maintenance facility. We have programmed the maintenance of all cross-references and can assure an accurate relationship between the data standards and the programming tables.

Finally, in the matter of electronic digital communications, there are really no technical problems. Parties can arrange to have their systems talk to one another through any of the existing common communications carrier facilities. TDCC performed a functional study and concluded that EDI required such features as distance insensitive, terminal insensitive, and speed insensitive services. A common communications protocol is highly desirable, along with link standards, and a form of distributed store and forward message handling. Communications carriers plan to introduce new services in the next few years to provide far greater ease in inter-system linkages.

In the past few minutes, I've given you a little theory, some principles, and some evidence that we have overcome the technical problems of EDI. We are now in the implementing stage. We are refining every day. The number of active interested companies is growing steadily but surely.

If I wanted to leave an impression with you it would be that industry is advancing its planning to provide for digital systems to replace paper methods as a means of transferring information between companies. When the critical mass of participants is reached, the change will be as rapid and as dramatic as the conversion from propeller to jet aircraft.

As you will recall, I stated at the beginning that my message would be brief, modest, and humble. I noticed a titter of laughter at that point and you were right. Thank you.

General Concept of a Data Resource Directory System

C. E. Hankerson

Department of Transportation
Federal Aviation Administration
Washington, D.C.

1. Introduction

It is a pleasure to be with you today and to have the opportunity to discuss with you some general concepts of a Data Resource Directory System (DRDS).

If you accept the view that data is a resource then you will accept the notion that systems can be designed to record and maintain these resources.

Figure 1. There are often data resource usage inefficiencies among the functions performed by an organization. If these inefficiencies are to be recognized and controlled, management must eventually inventory its total data resources to evaluate the overall effectiveness of these resources.

A Data Resource Directory (DRD), and its associated system is the tool by which efficient Data Resource Management can be achieved. It provides a means for collecting and identifying data resources.

Figure 2. The concept of a DRD and its use is designed for all users of data resources. This includes not only Data Resource Managers, but other users in areas such as Program Management, Records Management, Forms Management, and Data Processing Management.

2. Framework of a DRD

A person might ask what is the framework over which a DRD will be built? What gives it shape?

Figure 3. Well, let me develop a framework for you. If an organization is examined one may find that its purpose and size dictate a hierarchy of distinct functions. Each function is usually divided into subfunctions which are defined in increasing detail. This division continues until the organization has defined its operations at the lowest meaningful level. The functional division can now be hierarchically structured as shown here.

3. Operational Charactistics

Figure 4. Each function can be described in terms of three components; namely, Inputs, Processes, and Outputs. Inputs are the 'start-up' components on which Processes work. Processes makes possible the transformation of Inputs into Outputs. Outputs are the results of work performed on inputs by the Processes.

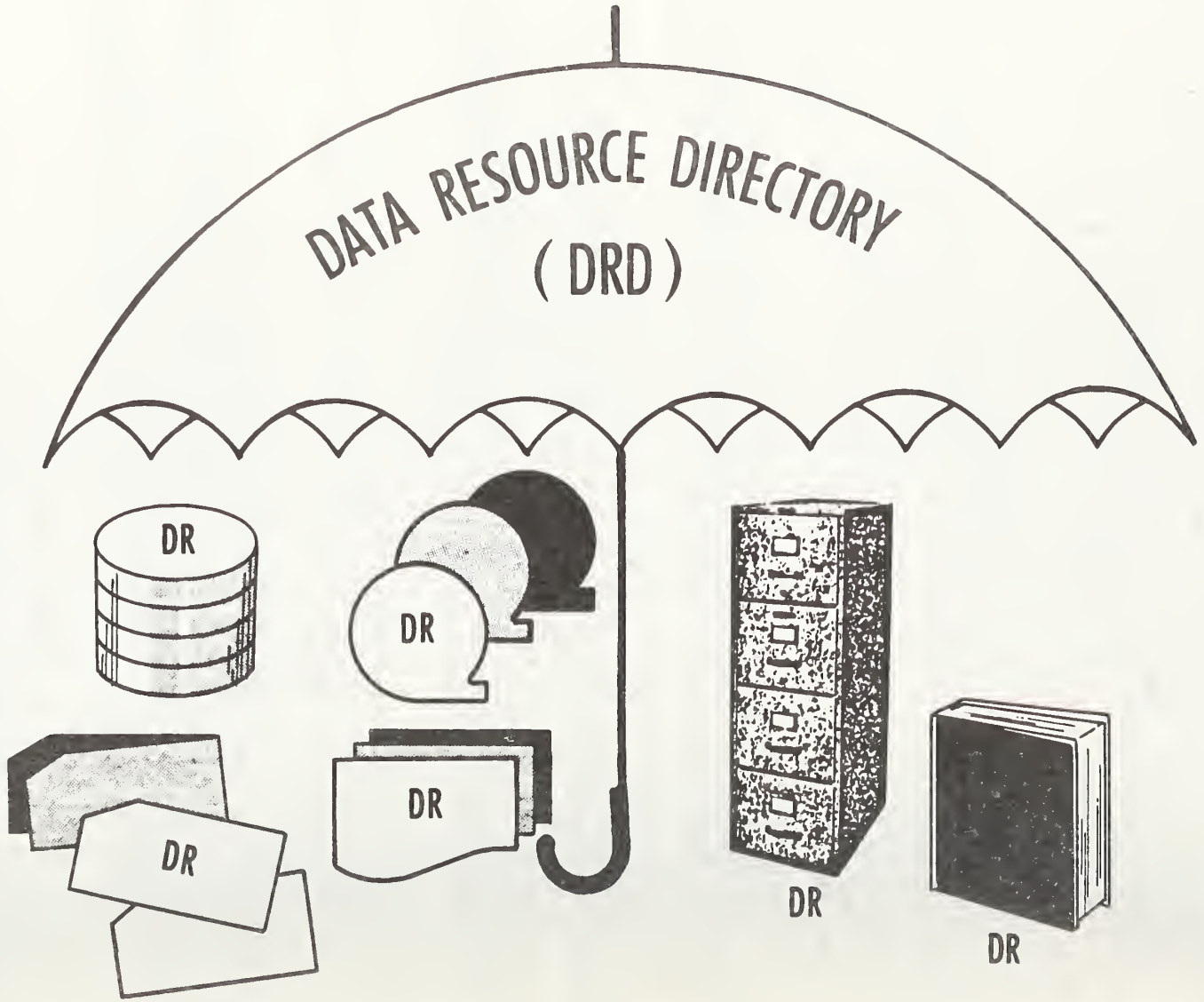


Figure 1

REPRESENTATIVE MANAGEMENT AREAS THAT WOULD USE A DRD

* Data Resource Management
* Program Management
* Reports Management
* Records Management
* Forms Management
* Data Processing Management

Figure 2

ORGANIZATIONAL FUNCTIONAL DIAGRAM

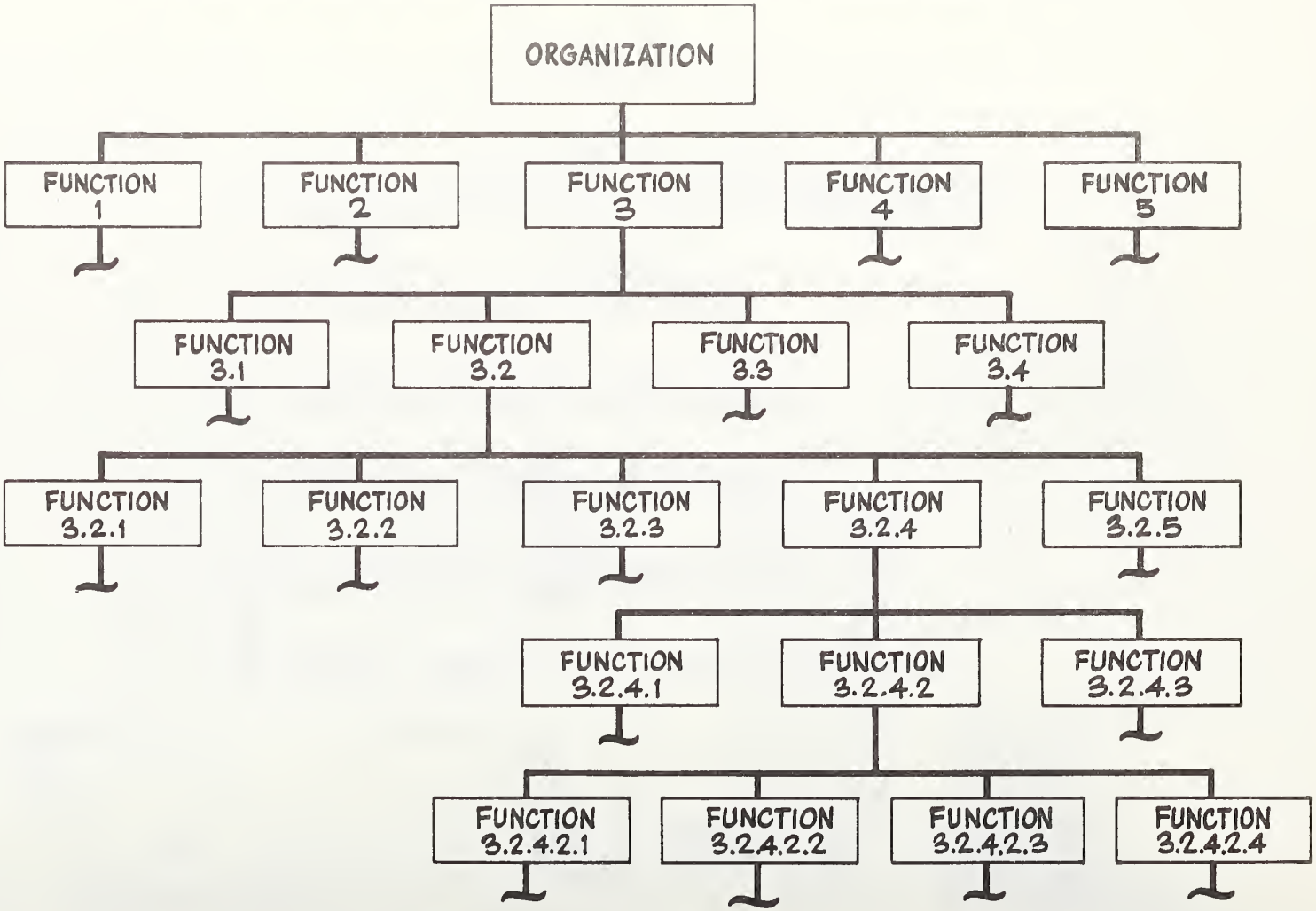
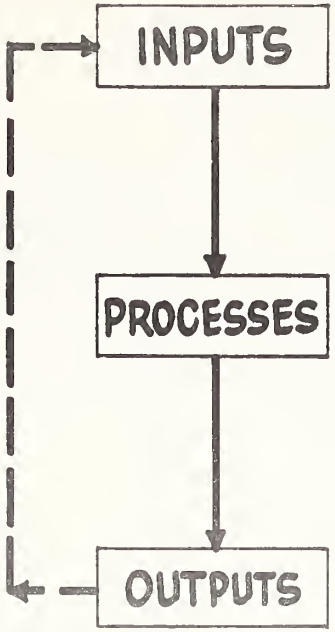


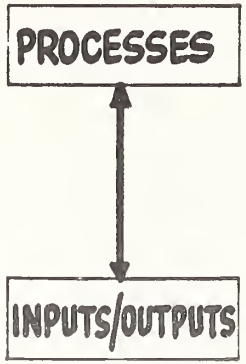
Figure 3

COMPONENT FUNCTIONAL HIERARCHY



A

ALTERNATE VIEW



B

Figure 4

Since most Inputs were once Outputs (i.e., The Inputs to one Process are usually the Outputs from another Process) these may single component; namely, The Input/Output component. In this view of functional components, entities classified as Processes work upon entities classified as Inputs/Outputs.

Figure 5. For the purpose of this presentation, four Entity Classes have been categorized as Processes: PLAN/PROGRAM, SYSTEM, APPLICATION, and PROCEDURE. Plans and Programs are comprised of Systems which are composed of Applications which are structured from Procedures. This hierarchy works on Entity Classes categorized as Inputs/Outputs.

Figure 6. For the purpose of this presentation five Entity Classes have been categorized as Inputs/Outputs: FORM, REPORT, FILE, RECORD, and DATA ELEMENT. Processes work on Forms, Reports, and Files; each of which may be comprised of Records, and all of which are composed of Data Elements.

Figure 7. When all Entity Classes are combined, we have a component functional hierarchy that looks like this.

The purpose of structuring this framework in the manner that is shown here, is for you to view the entities as manageable data resources.

You should note that what constitutes an Entity Class is organizational dependent. What one organization may decide is necessary may differ from what another considers necessary.

4. Data Resource Directory

It is now appropriate to offer a definition for a DRD. A DRD is defined as a centralized repository of the inventoried data resources of an organization.

It is important that a distinction be made between a DRD and other tools that are used in the management and control of data. The DRD approach is that all data (manual and automated) are resources, and to better manage this resource pertinent information about it should be readily accessible to all echelons of an organization. Other approaches (i.e., Data Element Dictionaries) are mainly used in the management of computer processed data elements.

Figure 8. In order for a DRD to be effective it must, as a minimum, serve the needs of managers in the following areas:

ACQUISITION	REDUCTION
ANALYSIS	RELATIONSHIP
DEFINITION	SECURITY
DESCRIPTION	SELECTION
IDENTIFICATION	SOURCE
LOCATION	STANDARDIZATION
PLANNING	USAGE

5. Major Components of a DRD

The major components of a DRD are Entities and Attributes. Entities are data resources that have distinct existence and definition, and are of

COMPONENT FUNCTIONAL HIERARCHY (PROCESS VIEW)

Figure 5

COMPONENT FUNCTIONAL HIERARCHY (INPUTS/OUTPUTS VIEW)

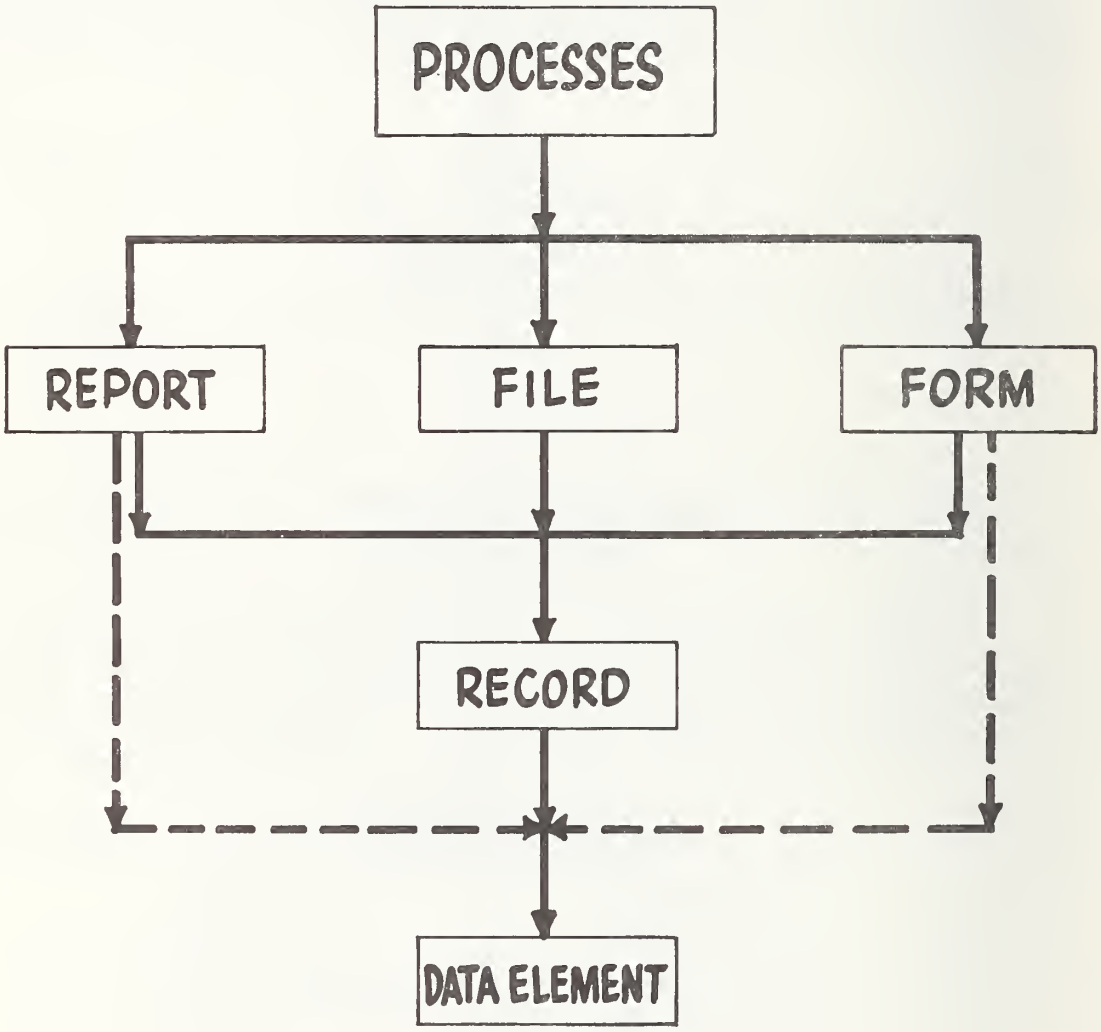


Figure 6

COMPONENT FUNCTIONAL HIERARCHY (ENTITY CLASS VIEW)

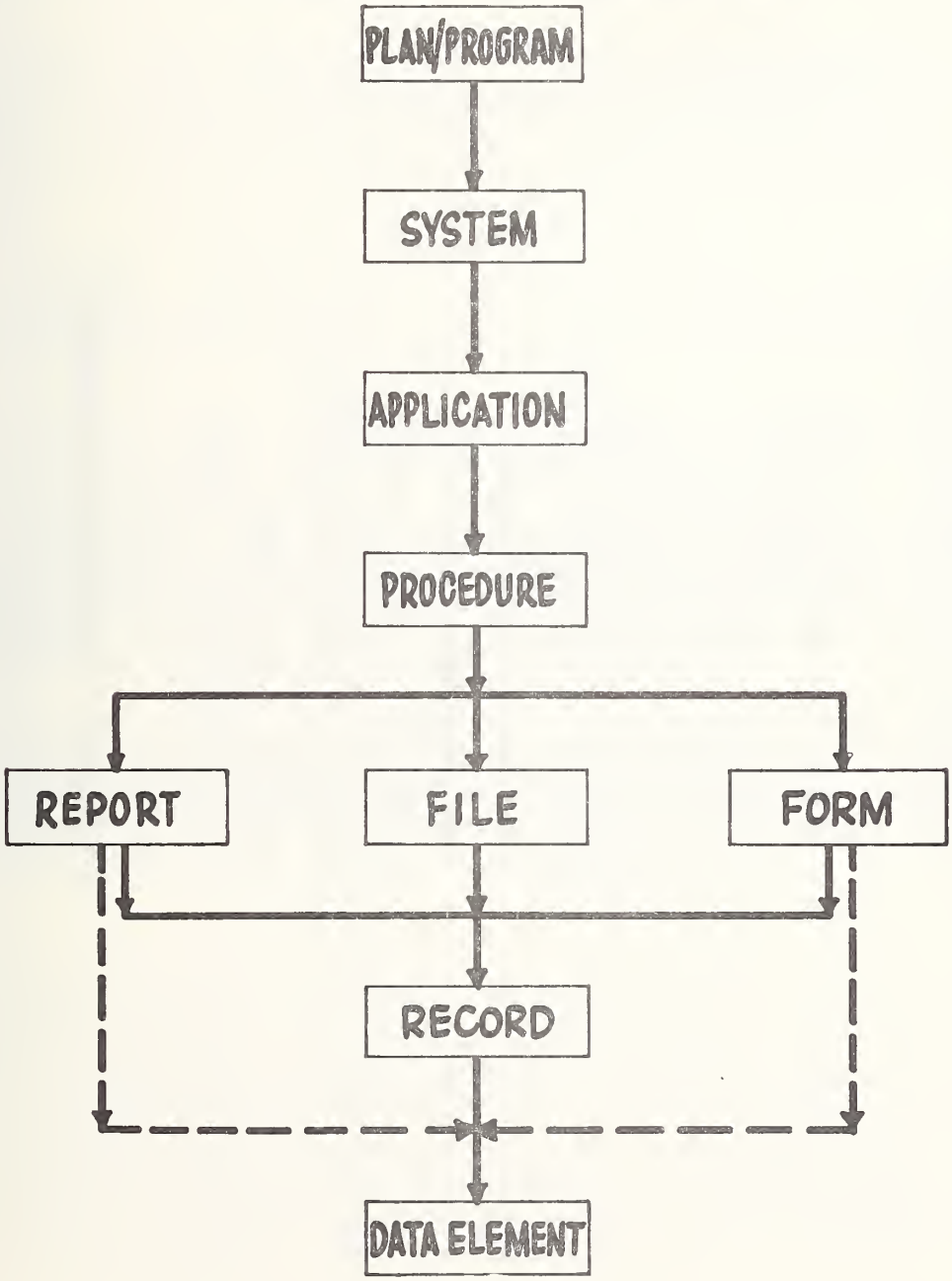


Figure 7

A DRD AIDS IN THE FOLLOWING AREAS OF ENDEAVOR

- DATA ACQUISITION
- DATA ANALYSIS
- DATA DEFINITION
- DATA DESCRIPTION
- DATA IDENTIFICATION
- DATA LOCATION
- PLANNING FOR DATA
- DATA REDUCTION
- DATA RELATIONSHIPS
- DATA SECURITY
- SELECTION OF DATA
- DATA SOURCES
- DATA STANDARDIZATION
- DATA USAGE

Figure 8

sufficient importance to an organization. Attributes are the characteristics or properties of an entity that define its context.

Several entities may support each function of an organization. The effective management of these data resources during their concept, design, development, and usage phases can best be accomplished by their inclusion in a central referrible repository. This repository can also be used to bridge the information gap between different organizations, and functions.

A DRD can be implemented on an organization-wide basis or at any functional level of an organization. It can be implemented one entity Class at a time, a combination of two or more Entity Classes, or all Entity Classes that an organization may decide to include in its DRD (An Entity Class is a group of entities having common characteristics).

6. Attributes of a DRD

Figure 9. Each entity has inherent attributes. The attributes recorded in a DRD relative to each entity are its contents. The attributes recorded reflect the information needs to effectively manage data resources.

This list is representative of attributes usually applicable to each of the entities. Notice that the attributes are listed in terms of attribute classes.

Attributes classes and attributes may differ among organizations. It is not the intent of TG-17 to identify all possible attributes that can be recorded in a DRD. Other attributes such as character type, format, scale, number of copies, type of form, benefits, etc., may be required. In general, if the attribute data is not used for decision making, then it should not be included in the DRD.

Figure 10. The attribute information contained in a DRD is usually a portion of the information that an organization maintain about its data resources. One organization may record more of its attribute information than another. This figure is an illustration of a representative ratio of the attributes that an organization may elect to include in its DRD.

7. Relationships Between Entities

Figure 11. The attribute class called Relationships is extremely important. Entities are related at various levels in a hierarchy of functions. They may assume additional characteristics applicable to these relationships. Relationships fall into two categories; inter-relationships and intra-relationships. Inter-relationships indicate the association between different Entity-Classes. These associative relationships may be categorized as superordinate commensurate, or subordinate. Intra-relationships indicate the association between occurrence of entities of the same Entity Class. For example, a report entity may be associated with another report if both are produced by the same procedures.

REPRESENTATIVE ATTRIBUTES FOR EACH ENTITY CLASS GROUPED BY ATTRIBUTE CLASS . . .

IDENTIFIERS
- ENTITY CLASS
- OFFICIAL NAME
- ABBREVIATION(S)
- SYNONYM(S)
- REFERENCE DESIGNATOR(S)
- ENTITY IDENTIFIER

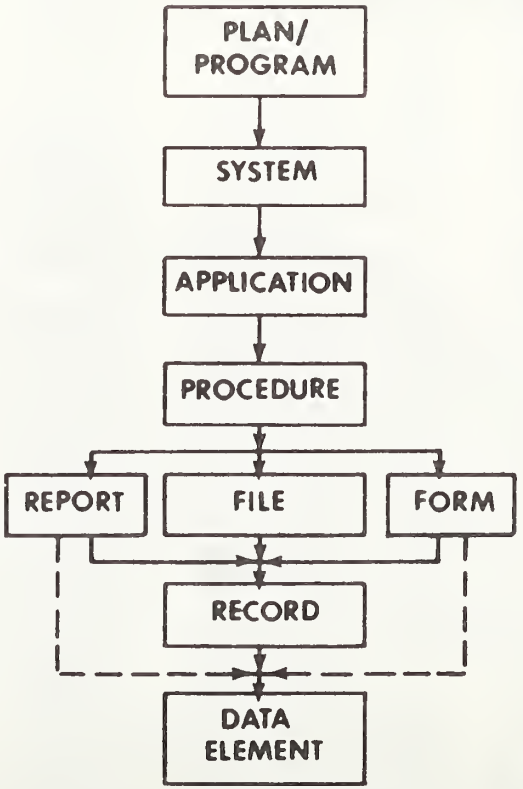
DESCRIPTION
- SEQUENCE
- SIZE
- PURPOSE
- SCOPE
- GENERAL DESCRIPTION

ORGANIZATIONAL RESPONSIBILITY
- PRIMARY RESPONSIBILITY
- SECONDARY RESPONSIBILITY

USER
- PRIMARY USER
- SECONDARY USER

REFERENCE DOCUMENTATION
- REFERENCE DOCUMENTATION

STANDARDIZATION
- LEVEL OF STANDARDIZATION
- STANDARDIZATION AUTHORITY



COMPONENT FUNCTIONAL HIERARCHY

(ENTITY CLASS VIEW)

MANAGEMENT CONTROLS
- DATE PROPOSED
- DATE DEVELOPED
- OPERATIONAL DATE
- LAST DATE REVISED
- DATE TERMINATED
- DUE DATE
- EXPIRATION DATE
- REVIEW DATE
- CUT-OFF DATE
- PERSONNEL COSTS
- EQUIPMENT COSTS
- SECURITY CLASSIFICATION
- INTER-AGENCY APPROVAL/CLEARANCE
- INTRA-AGENCY APPROVAL/CLEARANCE
- PRODUCTION FREQUENCY
- UPDATE FREQUENCY
- RETENTION PERIOD
- DISPOSAL DATE
- RETENTION AUTHORITY

RELATIONSHIPS
- ENTITY IDENTIFIER OF RELATED ENTITY
- OFFICIAL NAME/REFERENCE DESIGNATOR OF RELATED ENTITIES

Figure 9

REPRESENTATIVE RATIO OF DRD ATTRIBUTE CONTENT TO TOTAL ATTRIBUTES DOCUMENTED FOR EACH ENTITY CLASS

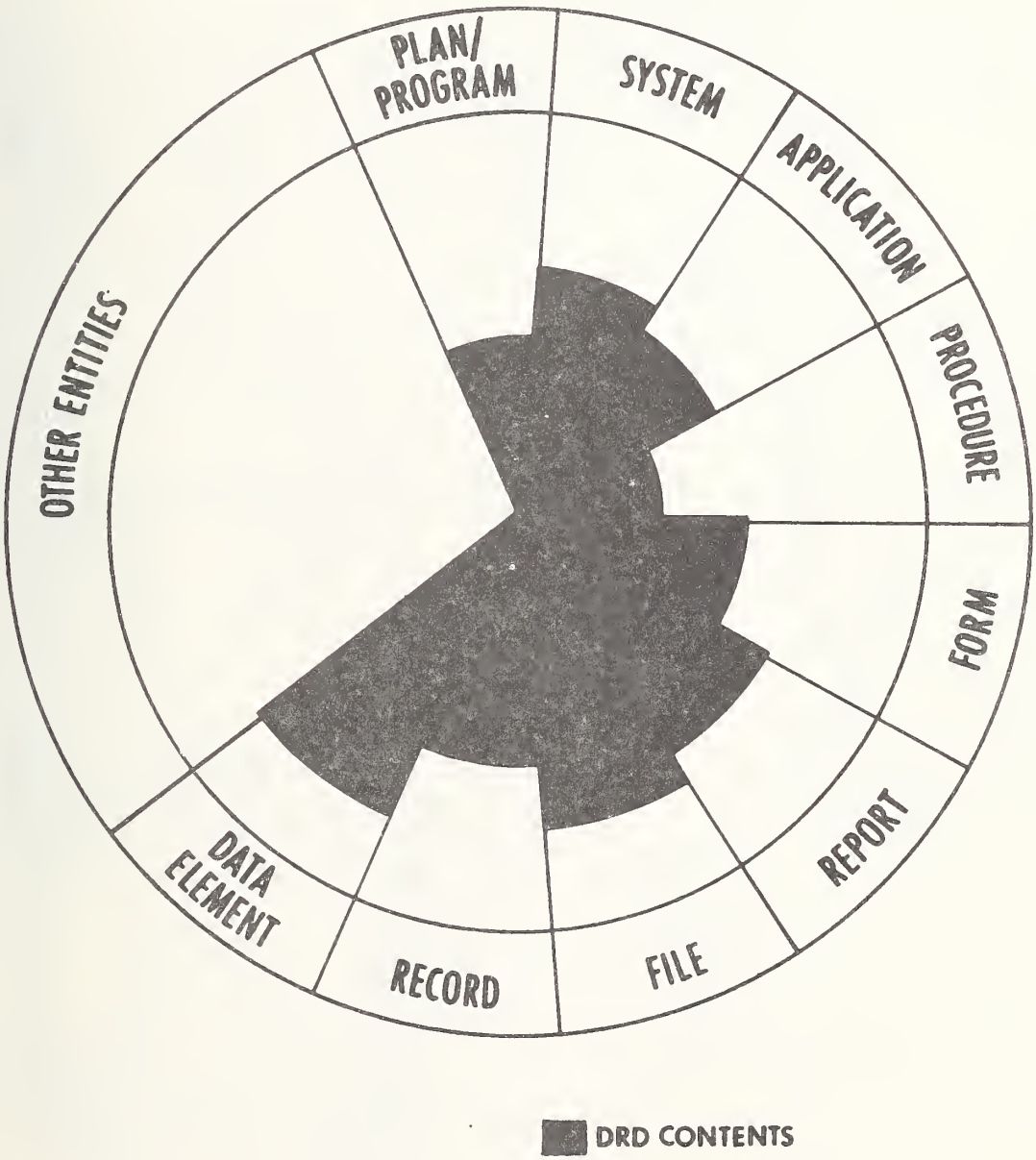


Figure 10

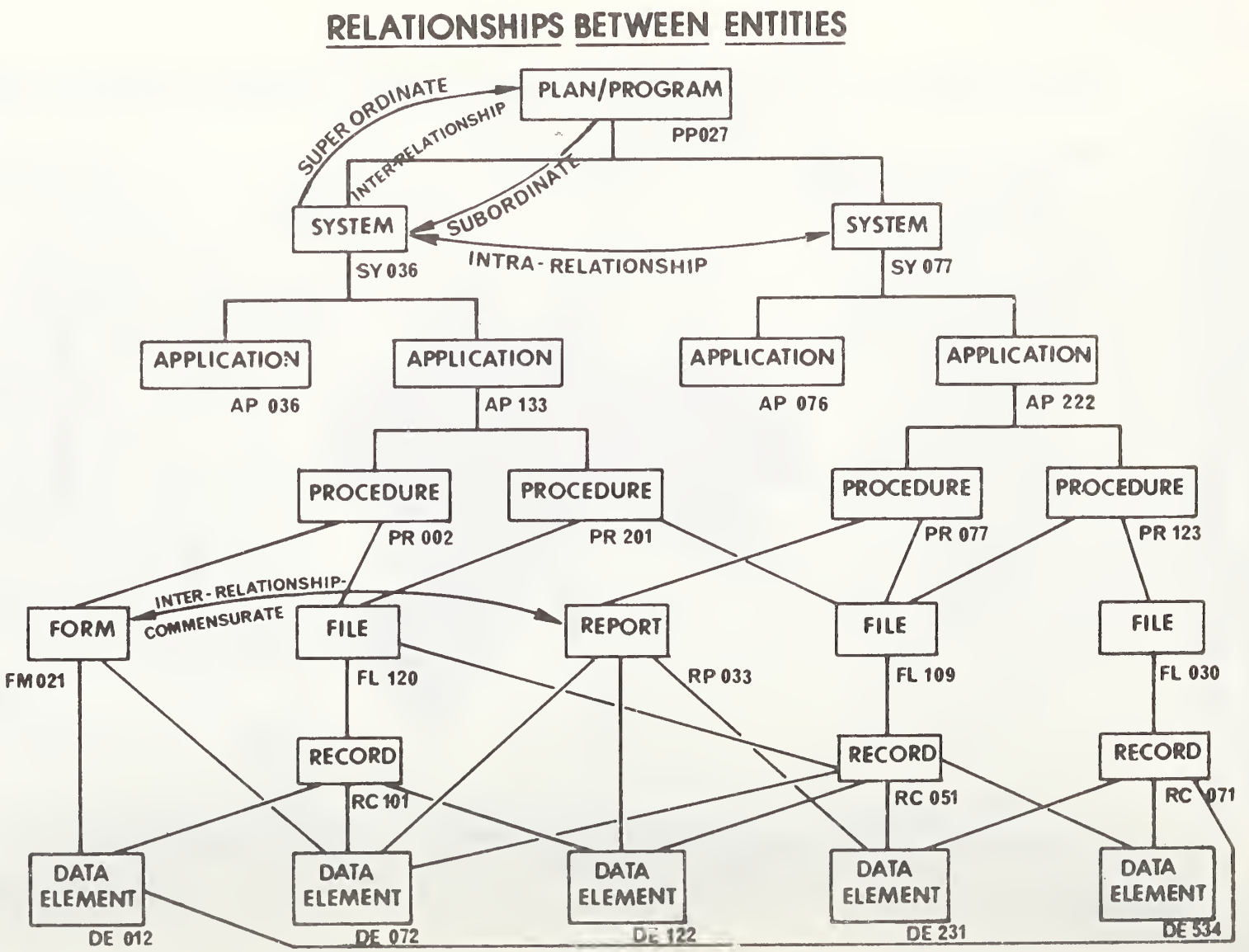

RELATIONSHIPS BETWEEN ENTITIES
PLAN/PROGRAM
PP027
SUPER ORDINATE
INTER-RELATIONSHIP
SUBORDINATE
SYSTEM
SY 036
INTRA-RELATIONSHIP
SYSTEM
SY 077
APPLICATION
AP 036
APPLICATION
AP 133
APPLICATION
AP 076
APPLICATION
AP 222
PROCEDURE
PR 002
PROCEDURE
PR 201
PROCEDURE
PR 077
PROCEDURE
PR 123
INTER-RELATIONSHIP-
COMMENSURATE
FORM
FM 021
FILE
FL 120
REPORT
RP 033
FILE
FL 109
FILE
FL 030
RECORD
RC 101
RECORD
RC 051
RECORD
RC 071
DATA ELEMENT
DE 012
DATA ELEMENT
DE 072
DATA ELEMENT
DE 122
DATA ELEMENT
DE 231
DATA ELEMENT
DE 534

Figure 11

8. Establishing Relationships

The importance of establishing relationships cannot be over emphasized. The implementor of a DRD has almost boundless options. Regardless of the option chosen, upon identification of an entity, a determination should be made whether or not it has related entities and the extent of the relationship. The entity identifier attribute is used to record these relationships.

When all relationships are recorded, along with all other entity attributes, a network will have been established that will enable one to retrieve all data of manageable interest. With this network managers could perform a review of new data or information requirements before the requirements are formalized or become de facto collection instruments.

9. Capabilities Provided by a DRD

In order for a DRD to be an effective tool it must be capable of providing information that aids managers in identifying, acquiring, disseminating, and controlling their data resources. The recording of the attributes of each entity in a standard manner provides the means for appraising the relative importance of each resource. The utility of this approach makes available the following list of capabilities:

a. The capability for determining the impact of anticipated, planned or approved changes in systems and implementing such changes.

b. What applications comprise a specific system?

c. What systems comprise a plan/program?

d. What procedures comprise an application?

e. The reports produced by a specific procedure.

f. The data elements contained in a report.

g. What procedure works upon a specific form?

h. What procedure works upon a specific file?

i. What files are worked upon by a specific procedure?

j. Enhance data standardization

k. What data element comprise a specific record?

l. The inter-relationships among systems.

m. What form is required for a report?

10. Goals and Objectives of a DRD

Figure 12. The goals and objectives of a DRD are:

a. Strengthening the management of data resources at every functional level of an organization by centralizing the data inventory process, and through reduction of unnecessary redundancy, unplanned inconsistency and inconsequental data resources.

b. Reduce cost of data resources at each stage throughout the life cycle of the resource: discovering, collecting, recording, handling, transferring, storing, retrieving, displaying, publishing, disseminating and disposing.

c. Strengthen data resources, and information about data resources, to meet the needs of decision-makers, researchers and problem solvers at all levels of an organization, with less costly data and data resource alternatives.

d. Permit more efficient determination of the impact of anticipated, proposed and/or approved changes by those organizational functions, planning, administering and maintaining plans and programs, and systems.

e. Establishing appropriate monitors, controls and surveillance measures to track the progression of a plan or actual activity.

f. Provide and maintain a central repository of data resources that relates to each function of an organization to service the entire organization.

g. The ultimate aim is to promote the development of multi-purpose, or common and shared use of data resources and discourage the proliferation of narrowly defined data for unique or special purposes.

11. Functional Requirements of a Data Resource Directory System (DRDS)

A Data Resource Directory System must provide the processes that allow the DRD to effectively function as a repository of data resources, and to be responsive to the needs of a wide variety sources and users of data resources. The basic system must allow for the acceptance, validation, storage and retrieval of many categories of data. The system must provide for substantial quality control of the directory, minimizing the use of human activity in areas prone to errors in directory input or output. Provisions must be made for DRD growth. Requirements also exist for an effective method of managing the development and operations of the system.

Figure 13. The total requirements of the system depends upon the level of implementation, complexity of the data, and the kinds of users and their expectations. As a minimum the DRDS must provide for:

a. Data capturing

b. Data validating

c. DRD updating

GOALS AND OBJECTIVES OF A DRD

- Strengthen the management of data resources
- Reduce cost of data resources
- Strengthen data resources, and information about data resources
- Determine the impact of anticipated, proposed and/or approved changes
- Track the progression of a plan or actual activity
- Provide a central repository of data resources
- Promote the common use of data resources

Figure 12

PROCESSES WITHIN A DATA RESOURCE DIRECTORY SYSTEM (DRDS)

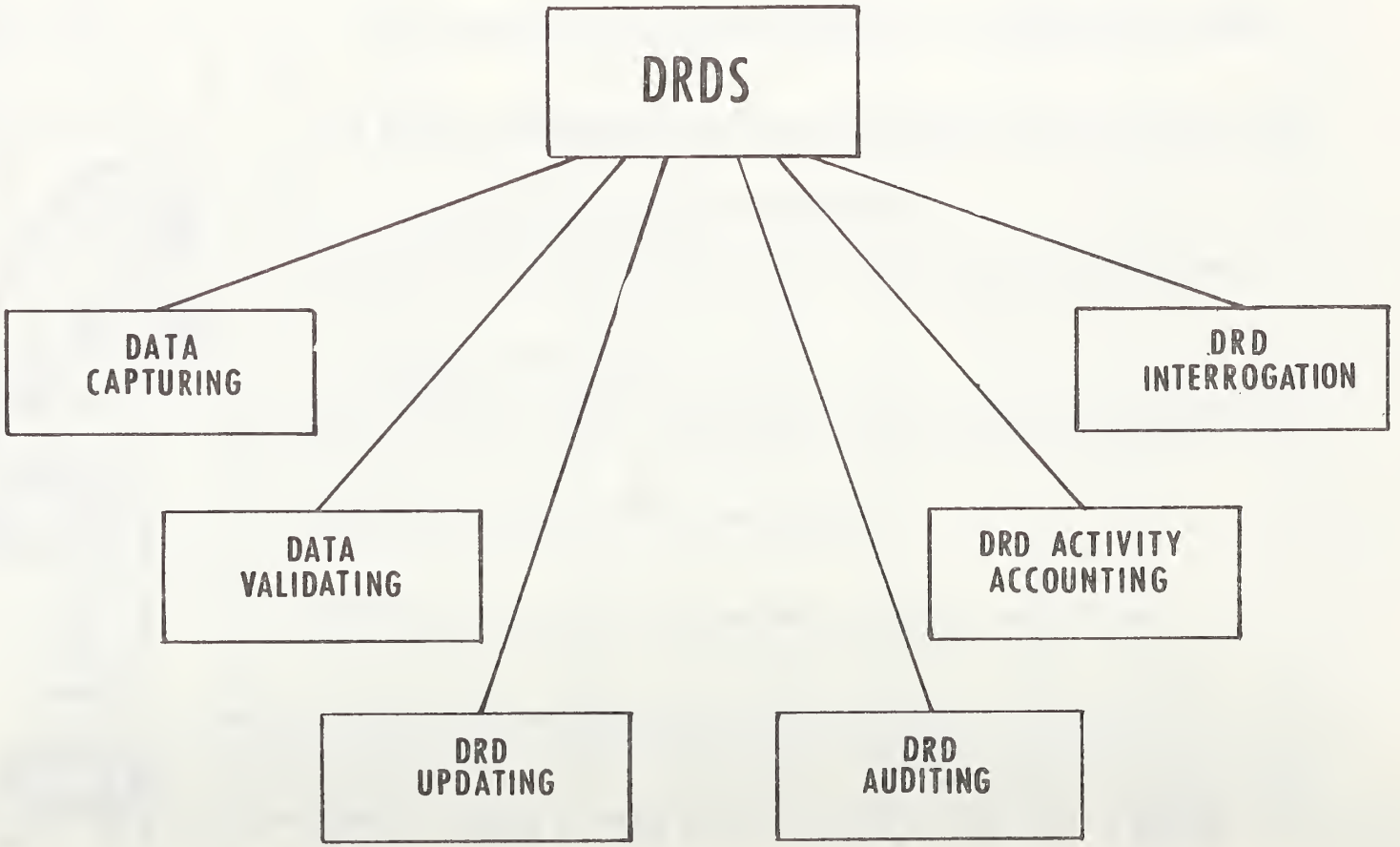


Figure 13

d. DRD auditing

e. DRD activity accounting

f. Interrogating the contents of the DRD

There are other important considerations such as utility functions. These include retrieving information from the directory, supporting directory maintenance activities, and the function of maintaining the integrity of the directory.

12. Summary

Careful planning for a DRD is a must. Policy and programs must be established to ensure that effective procedures are developed and that the DRD enables the organization to function in a coordinated manner and that it interacts within the organizational environment.

Any implementation effort should be guided by a plan that is designed to accomplish a specific set of Data Resource Management objectives.

A DRDS is not a panacea. It merely provides a more organized access to essential information upon which sound and effective decisions can be based.

International Standards Activities of the Economic Commission for Europe

Eugene A. Hemley[1]

National Committee on International
Trade Documentation (NCITD)
30 East 42nd Street

The United Nations Economic Commission for Europe has a program for the improvement of documentation and related procedures which impede the conduct of international trade. This effort has been assigned to a Working Party on Facilitation of International Trade Procedures which has divided its work under a Group of Experts on Automatic Data Processing and Coding (GE1) and a Group of Experts on Data Requirements and Documentation (GE2) Standardization work at present under GE1 is addressed to subjects such as Data Element and Codes, Trade Data Interchange, and Aligned Documentation and ADP. Under GE2, subjects being covered include a Unique Consignment Reference Number, Legal Questions and Problems, Import Documents, Dangerous Goods Documentation, and Definitions of Documentary Functions. Participation includes most European countries, Canada, Japan, Australia, and the United States, and the thirteen public and private international organizations.

Key Words: Alignment;dangerous goods; automatic data processing; common access reference number; data requirements; Department of Transportation, Office of Facilitation; documentation standardization; Economic Commission for Europe; import documents; export documents; legal questions; National Committee on International Trade Documentation (NCITD); United Nations; Working Party on Facilitation of International Trade Procedures.

1. Introduction

Although I am the Cardis Program Manager for the National Committee on International Trade Documentation and can supplement during the question period any of the information just given by Bob Cavanaugh on Cardis, my role at this point is to tell you about the role of NCITD in relation to the

[1]CARDIS Program Manager, Cargo Data Interchange System.

standards activities of the United Nations Economic Commission for Europe. This presentation should be given by Mr. Arthur E. Baylis, the Executive Director of NCITD who has been an advisor to the Department of Transportation the official United States representative for the ECE meetings for the past eight years, but unfortunately this meeting conflicted with the ECE meetings, and Mr. Baylis is now in Europe at those sessions. Since the Bureau of Standards work with ANSI and the International Standards Organization can better be covered by others at this meeting, I will concentrate my remarks on the ECE work in the standards area.

2. Organization

The ECE program for the simplification, standardization, and reduction of documentation and related procedural requirements that impede the conduct of international trade has been assigned to a Working Party on Facilitation of International Trade Procedures. Because of the diverse subjects which can be covered under this heading, the Working Party has divided its work into two major subjects which are handled by special Groups of Experts. These are:

Groups of Experts on Automatic Data Processing and Coding (GE1)
and
Group of Experts on Data Requirements and Documentation (GE2)

GE1, the former, covers the application of data processing to international trade transactions and GE2, the latter, concentrates on the simplification and standardization of applicable documentation and related procedures.

2.1 Task Teams

Even this division of the work was not sufficient and further organizational breakdown was found to be necessary for the specific topics being handled. These task teams with countries assigning chairmen are as follows:

GE1 ADP and Coding

1.1 Data Elements and Codes (Sweden)
1.2 Trade Data Interchange (U.K.)
1.3 Aligned Documentation and ADP (U.S.)

GE2 Data Requirements and Documentation

2.1 Unique Consignment Reference Number (France)
2.2 Legal Questions and Problems (International Chamber of Commerce)
2.3 Import Documents and Procedures (Czechoslovakia)
2.4 Alignment of Additional Trade Documents (France)
2.5 Dangerous Goods Documentation (U.K.)
2.6 Definitions of Documentary Functions (Canada & USSR)

In relation to this task team organization, it is anticipated that the work of Task Team 1.1 and 1.2 will be combined to address the inclusion of data elements in a message structure. With the work of the task team being scheduled for completion in 1978, it is contemplated that future assignments of this type will be given to individual rapporteurs, or chairmen, who will organize their own working sub-groups.

2.2 Organizations

Although this is primarily a European commission as noted by its title, the scope of its work has attracted the participation of many non-European countries such as Canada, Japan, Australia, and the U.S. Attendance is good and about 24 countries are regularly in attendance at the quarterly meetings in Geneva. However, because of travel expenses, it is anticipated that meetings in the future will be reduced to two a year. In addition, seven public and six private international organizations are also represented. These include:

Central Office for International Railway Transport
Council for Mutual Economic Assistance
Customs Cooperation Council
European Economic Community
General Agreement on Tariffs and Trade
Inter-Government Maritime Consultative Organization
United Nations Conference on Trade and Development

International Air Transport Association
International Chamber of Commerce
International Chamber of Shipping
International Organization for Standardization
International Rail Transport Committee
International Union of Railways

3. Coordination with International Organizations

An important function of this work is coordination with other international and national groups addressing the specified subjects. This vital coordination role is recognized by all participants as necessary to prevent duplication of effort or the creation of separately developed different standards on the same subjects. Parallel action which is being monitored includes:

Customs Cooperation Council (CCC)- work on development of a modernized, expanded Harmonized System of Commodity Descriptions and codes.

Inter-Government Maritime Consultative Organization (IMCO) - report on dangerous goods documentation and its 1977 Assembly which will specifically address the acceptance of electronically processed shipping documents.

International Association of Ports and Harbors - work related to international trade facilitation of which Mr. Baylis is one of the special advisors.

International Chamber of Commerce (ICC) - working party to study trade facilitation matters with the socialist countries. Mr. Baylis is Chairman of this working party.

Council for Mutual Economic Assistance (CMEA) - work on simplification and standardization of trade documents.

International Organization for Standardization (ISO) - work on maintaining country codes and money codes.

4. Subjects Under Study

The topics being actively addressed by the task teams include the following:

Currency codes - The merits of 2 alpha or 3 alpha codes are being explored and rationalized.

Country code - There still are several different codes in use and interest is shown in having a cross-reference matrix for the ECE 3 digit numerical code the UN numerical code, and the ISO 2 alpha and 3 alpha code.

Ships name codes - There are problems in maintaining the 60,000 ships radio call signs of the International Telecommunications Union which will be addressed in a special ITU Conference to overhaul the system. The Economic Commission for Latin America (ECLA) is considering a revision and an alternate system based on the ISO 2-Alpha country code.

Port Location Code - This work has been expanded to include not only seaports but airports, frontier crossing points, rail stations, and Customs ports.

Terms of Payment - The International Chamber of Commerce (ICC), and others, are submitting their versions of studies by Romania and Belguim.

Packaging Code - Various existing systems are being reviewed before preparing a consolidated recommendation.

Data Elements in Maritime Transport Documents - Extensive work on this subject has been done by the NCITD Cardis Committee who prepared a listing of the data elements used in the Bill of Lading and the Commercial Invoice with a breakdown by fields, elements, alternate names, definitions, and current coding. These lists were the basis of the U.S. submission to the ECE task team working on this subject.

Security Problems - Several countries has submitted papers on this subject which has been extensively studied by the NCITD Cardis Committees.

Trade Data Interchange Messages - Extensive work is proceeding on this subject also being studied in the Cardis report.

Unique Consignment Reference Number or Common Access Number - This project is devising a unique number by which shipments can be identified along their route from shipper to consignee.

Legal Problems - This subject which is reviewing possible legal problems in considering the transition from paper documents to data processing has also been extensively studied by the NCITD Cardis Committees. Possible problems relate to the requirements for signatures or other methods of authentication contract of carriage, negotiable instruments and evidence.

Import Documents and Procedures - This subject has received much attention in the United States. The U.S. Customs Service working with the advice of an NCITD committee has recently revised the Special Customs Invoice form, Customs Form 5515 to conform to the alignment standards of the U.S. Standard Master. Work is now proceeding on the revision of Customs Form 7501. This information is disseminated at the Import Documents and Procedures task team for use of countries importing to the United States and to request similar simplification in other countries.

Dangerous Goods and Documentation - This deals with the classification of dangerous goods and use of special data elements. NCITD has been doing extensive work in this area with industry and government, and has helped extensively in preparing the U.S. position. The varying approaches of carriers, shippers, and lawyers have been consolidated and a workable position has been reached. It is expected that the Task Team will complete its work this month and present its position to IMCO.

5. Conclusion

As you can see the work of the ECE related to international standards for international trade is heavily oriented towards documentation and data elements used in information processing. With United States industry's efforts proceeding at full speed towards simplifying international trade documentation and implementating a Cargo Data Interchange System, it is most important that the work of the ECE in this area not only be carefully monitored but that U.S. positions be strongly represented.

THE FEDERAL INFORMATION LOCATOR SYSTEM*

FOREST W. HORTON, JR. - STUDY DIRECTOR

Summary

Unprecedented Federal demands for information from the public and the severity of duplication in these overlapping requirements emphasize the need for coordinated action to reduce these burdens. The lack of a systematic and on-going mechanism to identify duplication and other reporting burdens imposed on the public has, more than any single factor, hindered dealing with these problems. The Government cannot identify the information that is currently being collected, where it is located, or how new reporting requirements relate to information that is already available. Effective solutions cannot be applied to a problem not adequately identified.

The Commission recommends the development of an inventory of these public reporting requirements. The inventory would be a single, authoritative register of all Federal reporting imposed on the public. Much like a catalog or index used in libraries, subject terms describing the general contents of these reports would be used to:

- identify duplication in existing or new reporting requirements;
- locate existing information that may meet the needs of an agency and thereby promote sharing to avoid duplication;
- provide a central coordinating mechanism for Federal, State and local government requirements for information;
- maximize the use of information by identifying available information for Congress in drafting legislation and information for the executive branch in operating programs; and
- make visible public burdens from this reporting so that effective action can be applied to reduce these burdens.

This registration, inventory, and index mechanism is named for the function it will perform—the Federal Information Locator System. It is an essential tool for improving the performance of the Office of Management and Budget (OMB), the General Accounting Office (GAO) and the agencies under the Federal Reports Act.

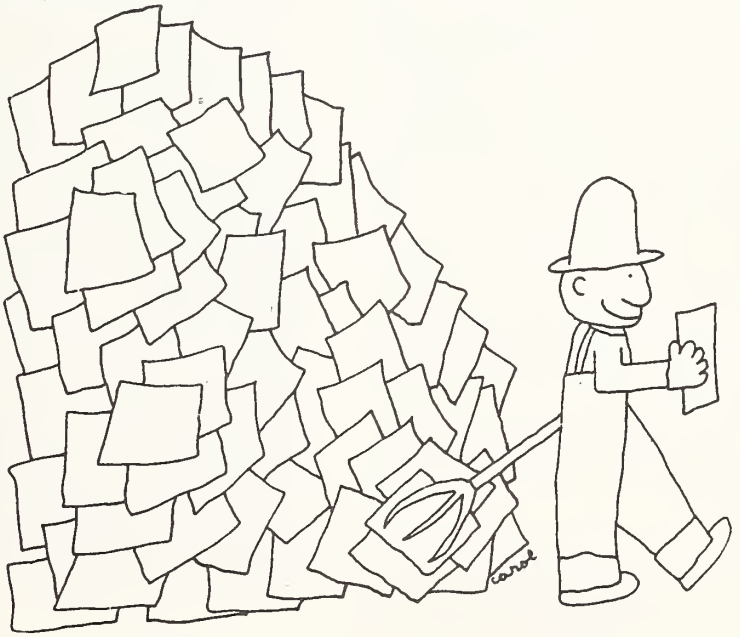
The locator system concept builds on capabilities already in existence, but scattered among many Federal agencies. A few States and local communities also have developed this approach. Effectiveness and simplicity in design and use will depend on a carefully phased, coordinated development plan, with pilot testing at each stage. Full participation and involvement of agencies, and other organizations in private industry and professional organizations, will be elicited. Their expertise and experiences will be drawn upon to the maximum extent.

The step-by-step evolution from concept to an operational, pilot-tested system will require a minimum of from three to five years. An estimated $1.2—2.4 million will be required to develop the prototype system based on the exact number of pilot test areas selected, and other considerations. The Commission has recommended that three or four major functional areas be selected for detailed investigation, such as energy, health, education, procurement, business financial information, and information on State and local government.

Additional funds would be required to extend the system, once developed and tested, Government-wide. Annual costs of $750,000 are projected to operate the system once it is developed and put in place.

These cost estimates are broad approximations based on an extension of the experiences and costs incurred by individual Government agencies and private organizations in similar efforts. Therefore, the cost estimates should be viewed as indicating the relative ranges for the amounts involved.

In view of planning and coordination needs, and the time required to achieve a working system, the Commission recommends that the Director of the Office of Management and Budget organize a task force to begin work toward the development of this important tool to reduce the paperwork and red tape burden on the American public.



*Commission on Federal Paperwork
July 15, 1977

I.
Introduction

Eliminating duplication and unnecessary differences in the information requirements imposed on citizens, businesses and others by government first requires knowledge of what information is already available within government, and what is planned to be collected. In short, an index or "locator system" is needed for this purpose. Once existing and planned data are identified, the nature and severity of duplication can be determined and steps then taken to control it.

The Commission documented in several of its studies that duplication and overlap in information collected from the public by multiple government agencies is a very serious problem. While eliminating duplication will require extensive and detailed work to settle the differences among program officials and agencies with similar data needs, what is missing to make the first step is a tool to help officials quickly and efficiently learn what information is already on-hand, and planned for collection.

The problem is, the Federal Government does not now know what information it collects, with what frequency, from whom, and for what uses. A single, comprehensive, and up-to-date inventory does not exist. The consequences of this "knowledge gap" are far-reaching and mostly negative.

Filling this gap requires a *Directory of Federal Information Resources* and a *Dictionary of Federal Data Elements*. The Directory would serve as an index to existing and planned information resources, and the Dictionary would serve as a central, authoritative compendium of standard definitions for common terms and abbreviations used throughout government.

Like all tools, the effectiveness of the Directory and Dictionary depends upon:

- how well these tools are designed to serve the purposes for which they are intended;
- how efficiently the tools are used; and
- whether or not incentives and sanctions are appropriately applied to insure that the tools are used when, and only when, required; and for the purposes, and only for the purposes for which they are intended.

In short, if management is not involved at each stage of development, testing, and implementation, or is lax in attitude toward enforcement of its use, it will be ineffective.

The Commission's Study of Duplication

The need for a Federal Information Locator was identified in the Commission's *Duplication* study. That study examined the reporting requirements of one important business sector, food chain stores. The Association of Food Chains established a paperwork task force to provide the Commission with an inventory of Federal reporting burdens placed on its members. The inventory identified 128 Federal forms and reports which must be submitted to Government agencies either voluntarily or mandatorily. Subsequently, the Association furnished a list of recommendations and complaints that included several cases of duplication. Forms identified in these complaints were used to initiate an analysis of duplication, including its significance, the manner in which it occurs, various types of duplication, and its severity. Additional forms were selected as potentially duplicatory based on the form titles, subjects covered, or requesting agencies. The analysis included forms from six agencies: Department of Labor (DOL); Internal Revenue Service (IRS); Federal Trade Commission (FTC); Securities and Exchange Commission (SEC); Department of Agriculture (USDA); and Department of Commerce (COM). Subject categories included: Business activity/Product; employees; ownership; property; legal; respondent organization description; securities; and finances.

Examination of these forms and reports revealed that "duplication" was not a unitary concept but one that requires three categories to differentiate between levels of likeness and severity:

- identical duplication;
- similar duplication; and
- generic duplication.

In degree of *likeness*, identical duplication is the most specific because the information requested is, by definition, exactly the same. However, in degree of *severity*, generic duplication, representing groups of related data, is the most severe.

Identical duplication occurs when two or more individual elements of data have the same definition or meaning. The nomenclature used to describe the two elements of data may differ, but if the definitions or meanings are the same, identical duplication exists.

Similar duplication involves individual data elements related to the same specific subject, but with minor differences in meaning. For example, if one report asks for the "dollar amount of a certain line of merchandise bought and resold at retail," and another asks for the "estimated percent of total sales," rather than dollar value for merchandise, similar duplication is considered to exist.

Generic duplication occurs where two or more reports request groups of data that relate to the same subject. Another way of expressing this is to say that similar duplication involves differences in definitions at the individual data element level whereas generic duplication involves dissimilarities in definitions of broad groupings of data. For example, in examining five major groupings of business financial data—gross sales, expenses, income, assets, and liabilities—all three duplication conditions were noted. In some cases individual data elements exactly duplicated one another—identical duplication. In others, individual data elements relating to the same major grouping (e.g., gross sales) were similar, but were defined slightly differently—similar duplication. And, in looking at several major groupings of data as a whole, such as expenses and income, there was significant disparity in how they were defined and treated as between two or more different reports—generic duplication.

The study revealed that duplication is substantial and does impose severe burdens on respondents. This severity can be expressed as a percentage of total data elements in a single report that are duplicative of those in another report. In all of the cases analyzed, the occurrences of duplication ranged in severity for the three levels of defined duplication as follows:

Identical duplication	10%	65%
Similar duplication	9%	50%
Generic duplication	29%	92%

These results indicate that the most severe form of duplication is generic. It represents the greatest burden on respondents because respondents must collect multiple sets of data and maintain multiple records for the same subject. Problems are often created in reconciling different sets of data and certifying their accuracy. In some instances, additional collection and reaggregation of data become so burdensome on respondents that estimates are submitted rather than data representing actual situations.

Identical duplication presents relatively less of a problem to respondents because once an element of data is collected, compiled, and available to the respondent, it can be furnished in identical form to as many requesters as required. Substantial burdens are presented when respondents must separately count, tabulate, and maintain data to reflect differing time periods or slightly different aspects of the same subject. This problem is prevalent in both generic and similar duplication. Thus, reduction of similar and generic duplication can produce greater relief in respondent burdens than will the reduction of identical duplication.

Reducing Duplication Requires Knowledge of What Information Exists

The Locator System has two parts. The first is the directory of reports and other information. This would be the index component which identifies what types of information are already on-hand and planned, with references to where they are located. The second component is the dictionary of standard terms, or data element dictionary as it is conventionally called. Common use terms and their standard definitions would be alphabetically listed, cross-referencing the use of the term to agency, information system, and report form. One of the principal uses of the directory will be to identify duplication in both existing and planned reporting requirements. The data element dictionary will be used to help resolve differences in definitions for similar data elements.

The Information Directory

The problem in identifying duplication without a systematic mechanism can be illustrated by the magnitude of Federal public use reporting. There are approximately 10,000 different Federal public use reports, requesting close to 1.2 million elements of data from the public. Lacking a mechanism to track and manage such vast quantities of report data, the total size and character of duplication cannot be measured, nor can all cases be identified, nor (most importantly) can future duplication be avoided.

Controlling Future Duplication. For paperwork and information duplication to be controlled and reduced it must be caught in the early information planning stages. The locator system would serve as a planning activity across agency lines and thereby encourage sharing and illuminate significant gaps and redundancies in Government's information. More specifically, the use of a locator system as a planning tool could:

- Assist agencies in planning, coordinating, and evaluating their information requirements more efficiently and effectively since they would be aware of what already exists or that which is planned to be collected;
- Maximize the use of the data and information already available in the Federal establishment, thereby avoiding the unnecessary recollection of duplicative data from the public, and the establishing of redundant and costly new information flows;
- Assist agencies in fulfilling their responsibilities under the Freedom of Information Act and the Privacy Act, by serving as an authoritative, consolidated master reference index of agency data holdings and record systems, properly indentified as (1) open to public access or (2) not to be disclosed, as the case may be;
- Help detect, identify and root out generic, similar and identical data element duplication in information flows and individual documents (such as public-use reports), and thereby reduce and control existing as well as potential duplication and overlap;
- Serve as a research and analytical instrument to help detect, identify, and correct gaps in existing data, document and literature holdings, where agency information requirements are unfulfilled, partially fulfilled, or inefficiently fulfilled; and
- Serve as an authoritative, centralized reference and finding aid to assist both citizens and organizations within the private sector, as well as officials and others within the Federal establishment, to identify the existence of, and locate efficiently and accurately, data, document, and literature holdings within the Federal establishment.

But again it must be emphasized that the usefulness of the locator as a tool is largely dependent on the extent to which management supports its use and its enforcement.

Tool to Support The Reports Clearance Process. The Commission's *Clearance, Statistics,* and *Information Resources Management* studies all pointed to the need for a comprehensive index of Federal data and document holdings if the clearance process is to operate effectively and efficiently to control the proliferation of public-use reports. The locator mechanism would register, inventory, identify, and index the subject contents and other characteristics of all public use reports, thereby filling a management knowledge gap that now prevents public use reporting problems from being effectively addressed.

Both agency level clearance offices and the central clearance authority in OMB could use the locator system to determine whether or not the information they plan to collect is already available, or whether existing data, while not identical, may serve the agency's needs and thereby preclude the collection of new data.

Duplication From Overlapping Program Responsibilities. An important cause of duplication is the extensive overlap between and among agencies involved in the administration of the same or similar programs. For example, the *1975 Catalog of Federal Domestic Assistance* contained 1,009 programs administered by 55 Federal agencies. The number of programs and agencies have continued to increase.

One means of ascertaining the character and magnitude of agency overlap in program administration is to examine the budget functional classification scheme. Functional classifications have been used for many years to display the President's Budget in a limited number of categories according to central purpose, regardless of administering agency. Under the new congressional budget process, functional classifications provide a framework in which the Congress can make and communicate its decisions allocating Federal resources among competing national priorities.

A few examples illustrate the potential for overlap and duplication in collecting information from the public. For example, Budget Function 306 deals with Federal energy programs. At least seven major agencies collect energy information: Agriculture, Environmental Protection Agency, Federal Energy Administration, Energy Research and Development Administration, Interior, Federal Power Commission, Commerce and the Nuclear Regulatory Commission. Many other agencies also collect energy data. The Commission's energy study extensively documents the nature and extent of overlap in the Federal energy information areas. This overlap is shown in Figure 1, which lists the Federal organizations[1] operationing 279 energy programs. Eighty-eight different data bases are operated by these agencies to support the programs. Information on coal is maintained by 36 agencies; information on natural gas by 37 of these agencies. These overlapping information collections were identified with the Federal Energy Information Locator System (FEILS) operated by the Federal Energy Administration. Based on these general identifications, the system further identifies specific cases of duplicate data in related collections. FEILS is one of several systems that demonstrates utility of the locator system concept.

In other examples, Budget Function 403 deals with advancement and regulation of commerce. Information in this area is collected by the Departments of Commerce, Agriculture, Interior, and Housing and Urban Development as well as by the General Services Administration, the Small Business Administration, the Commodity Futures Trading Commission and several independent regulatory commissions.

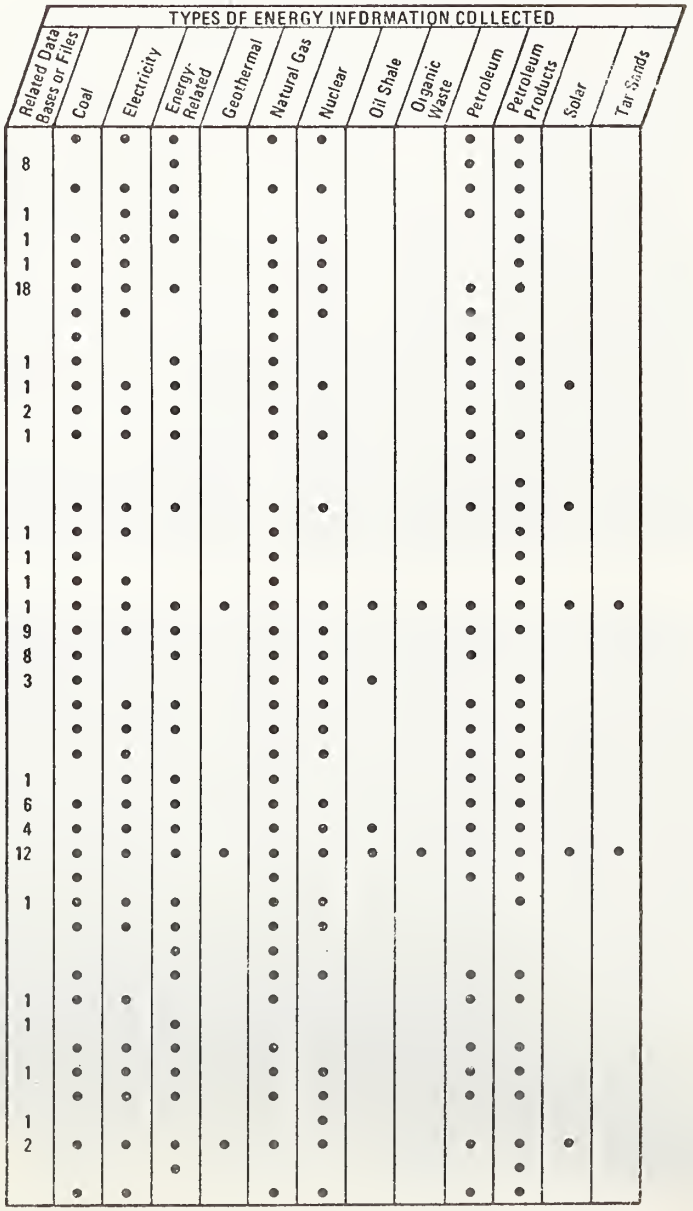
[1] A number of the organizations have been subsumed in the new Department of Energy.

FIGURE 1 PART 1.

OVERLAPPING ENERGY INFORMATION SOURCES

AGENCY NAME AND NUMBER OF PROGRAMS
Appalachian Regional Commission (10)
Civil Aeronautics Board, Bureau of Accounts and Statistics (2)
Central Intelligence Agency (1)
Department of Agriculture (6)
Department of Commerce, Bureau of Domestic Commerce (1)
Department of Commerce, Bureau of Economic Analysis (3)
Department of Commerce, Bureau of the Census (21)
Department of Commerce, Domestic and International Business Administration (1)
Department of Commerce, International Economic Policy and Research (1)
Department of Commerce, Maritime Administration (2)
Department of Commerce, National Bureau of Standards (4)
Department of Commerce, National Oceanic and Atmospheric Administration (2)
Department of Commerce, U. S. Patent Office (1)
Department of Commerce, U. S. Travel Service (1)
Department of Defense, Air Force (2)
Department of Defense, Defense Intelligence Agency (1)
Department of Defense, Defense Supply Agency (1)
Department of Defense, Navy (1)
Department of Defense, Office Assistant Secretary of Defense (1)
Department of Interior, Bureau of Land Management (4)
Department of Interior, Bureau of Mines (9)
Department of Interior, Geological Survey Conservation Division (5)
Department of Interior, Geological Survey Geologic Division (4)
Department of Justice, Economics Policy Office (1)
Department of Labor, Bureau of Labor Statistics (18)
Department of State (2)
Department of Transportation (9)
Environmental Protection Agency (5)
Energy Research and Development Administration (20)
Federal Energy Administration (38)
Federal Power Commission, Bureau of Natural Gas (19)
Federal Power Commission, Bureau of Power (8)
Federal Power Commission, Office of Accounting and Finance (13)
Federal Power Commission, Office of Economics (2)
Federal Trade Commission (6)
General Services Administration, Federal Preparedness Agency (1)
General Services Administration, Federal Supply Service (4)
General Services Administration, Public Buildings Service (4)
Department of Housing and Urban Development (4)
Interstate Commerce Commission (8)
Nuclear Regulatory Commission (2)
National Science Foundation (28)
Small Business Administration (1)
Securities and Exchange Commission (2)

FIGURE 1 PART 2.



Information related to community development programs (Budget and Function 451) is obtained by these Federal organizations: the Departments of Housing and Urban Development, Commerce, Agriculture, and by the Community Services Administration and ACTION. Here again, potential overlap is extensive. As a fourth example, Budget Function 741 (dealing with enforcement and prosecution) will include, at a minimum, the Departments of Justice; Health, Education, and Welfare; Treasury; the Equal Employment Opportunity Commission; and many regulatory commissions.

This kind of listing could go on and on. Because no comprehensive cross-agency and cross-functional inventory of existing and planned Federal information requirements has ever been undertaken, the full and true extent of overlap and duplication of information requests is unknown. An example of the direction that should be taken in identifying and cataloging information resources is the 1976 *Federal Information Sources and Systems* directory prepared by the General Accounting Office as part of the *Congressional Sourcebook* series. This directory describes recurring reports to Congress as well as executive agency systems containing fiscal, budgetary, and program related data.

The Data Element Dictionary

A locator system will help find similar information that may already be available, and where it is. After these data have been identified, the problem that must be resolved is use of the identified data for alternative purposes. The solution must center on meaning and usage.

A large number of existing data elements have the same name but different meanings among Federal agencies. This finding—documented in Commission studies on energy, public health, and welfare and financial reporting of companies—demonstrates the need for a Federal data element dictionary. The dictionary would facilitate the use or sharing of data for multiple purposes by providing:

- All meanings for commonly used words, terms, names, and abbreviations;
- A standard definition for each meaning; and
- Identification of agencies using which meanings for what purposes, and in which information systems, reports, and records containing the data.

Standard and uniform terms and definitions not only help insure and safeguard the integrity of data, but also provide us with a practical guide to help sort out meanings, users of data, and uses of data, by specific function centers. Data must not only be defined in terms of what it does represent, but also by what it does not represent.

Need for Information Processing Standards

Another prevalent problem relating to the lack of standard meanings for common use terms and names is the structure and manner in which data are represented. In manual systems or records the problem is relatively minimal. There, the human mind and eye can assimilate a variety of representations for the same information and compensate for the differences. However, these differences present a greater problem when data are compiled and exchanged using modern automation and information handling technologies.

For example, the lack of consistency and standards for citations shown on catalog card entries in libraries was one of the major factors inhibiting automation of an index to the vast literature holdings of the Library of Congress. Further, the exchange and assimilation of data where a State, for example, is represented as N. Car.; NC; or 18, becomes extremely difficult. And the representations of a single author as: John William Jones; and J. Jones, create separate entries in a bibliographic index as if these were different people. The differences create finding problems for users of an index.

Where a system cannot assimilate these differences to produce a single reference entry, this assimilation burden is placed on the users. Thus, the utility of information is substantially reduced in the absence of standard terms and definitions.

The Dictionary of Federal Data Elements will serve as a reference to standard terms and definitions as well as an enforcement tool for the application of these standards. It would be linked to the Directory of Federal Information Resources to identify the location of data elements in various Federal reports, records, and collections. Each standard will encompass nomenclature used to identify the element of data, its definitions, and technical specifications for representation of actual data values for the element.

The National Bureau of Standards, in its Federal Information Processing Series (FIPS), has already developed a guideline for documenting descriptive information for data elements. This guideline should be used as the basis for a Dictionary of Federal Data Elements. And the data standards already developed by the National Bureau of Standards, Institute for Computer Sciences and Technology, should be incorporated into the Dictionary.

In short, the Dictionary, like the Directory of Federal Information Resources, would provide a single, central, authoritative reference for Federal data collected and in use. The application of standard terms and definitions would increase the quality of data collected, and improve the utility of data already in use. It would also facilitate the sharing, interchange, and multiple use of existing data and thereby preclude duplicative collection of data already available. Central to all of these points is management and control.

Not a New Idea

Data element directories and dictionaries are not new or revolutionary concepts. Their use is widespread in Government agencies and private industry. A sampling of these locator systems examined by the Commission is shown in Appendix C. Although effectiveness varies from system to system, the basic conceptual, technical, and economic principles of these systems were found to be sound.

II.
How the Locator Would Work

The Locator is essentially a finding tool, using an index of subject terms or descriptors keyed to report contents. This type of index is substantially smaller and provides greater utility than an index containing discrete data element names.

In brief, the system would work as follows:

- Agencies planning to collect information from the public would be required to structure their requirements in terms of key characteristics (see Figure 2 for a list of suggested characteristics).
- Each of these characteristics, including the subject descriptors, for a proposed new report would be entered and registered in a centralized, automated indexing system as a "report profile".
- These profiles (whether for proposed new reports or revisions to existing reports) would be matched against existing profiles and would be made available to a central review office authority as well as the agency reports officer.
- Instances of identical duplication would be examined first. The burden of proof would be on the requesting agency to demonstrate why it could not use identical information already available.
- Next, similar duplication would be considered. (The utility and need for a Data Element Dictionary are evident here.)
- Third, generic duplication would be addressed, again involving the Data Element Dictionary.
- Proposed reports would be registered in the system and available for reference and coordination with any other requirements that may simultaneously be in the planning stages. Some evidence indicates that when crises such as the energy crisis occur, there is a "shotgun" effect whereby many agencies rush to collect the same or similar information simultaneously.
- After agency coordinations have been effected, and exchange, sharing, access, confidentiality and safeguard constraints identified, the planned requirement would be formalized for respondent coordination to determine:
  - — availability of the data;
  - — quality in terms of accuracy, completeness, timeliness, and relevance;
  - — undue burdens that may result from the planned requirement if implemented (for example, if the collection would impact a respondent or respondent category already overly burdened);
  - — suggestions for alternative data that would still satisfy the basic requirements, albeit in less than "perfect" ways (for example, sampling);
- A case file containing correspondence from all of the agencies and other parties, and documenting their positions and decisions, suggestions from respondents for simplifying and streamlining proposals, and a highlight summary of decisions/actions for top review authorities at both central and agency levels would be prepared.

Even though Commission studies demonstrate that the problem of duplication can be clearly seen in public use reporting programs, eliminating this duplication requires cooperation and extensive work to settle the differences among program officials and agencies with similar data needs. The Commission is proposing the designation of selected "focal agencies," or "centers," to coordinate the information collections of similar programs, on a continuing basis, such as energy, health, education, procurement, business financial information, and information on State and local government. In the first four areas, there is already extensive work being performed to develop an information locator capability. These centers would do much of the data collection, but would be principally responsible for planning, coordinating, and developing standards applicable to data collections in their fields and providing technical advice and assistance. Interagency *ad hoc* efforts to control and resolve duplication should be backed by OMB pressure to exhaust all practical alternatives for eliminating duplication.

FIGURE 2

PUBLIC USE REPORT PROFILE

1. SERIAL REFERENCE NUMBER
2. TITLE OF FORM OR REPORT
3. DESCRIPTIVE NOTATION OF CONTENT
4. AGENCY FORM OR REPORT NUMBER
5. DATE INSTITUTED, MONTH, YEAR
6. DATE TERMINATED
7. AGENCY NAME
8. ORGANIZATION NAME
9. PROGRAM NAME
10. PROGRAM OBJECTIVES: • Delivery of Benefits, Services or Asistance • Regulation or Enforcement • Licensing • Public Safety • National Defense • General Government Management and Administration • Revenue Collection
11. BUDGET FUNCTION CODE
12. ESTIMATED START-UP COST
13. ANNUAL OPERATING COST
14. FREQUENCY OF COLLECTION
15. NUMBER OF RESPONDENTS
16. AGENCY ESTIMATE OF RESPONDENT MAN-HOURS PER REPORT SUBMISSION
17. RESPONDENT ESTIMATES OF MAN-HOURS PER REPORT SUBMISSION
18. NUMBER OF DATA ELEMENTS ON FORM OR REPORT
19. ESTIMATED CHARACTER VOLUME PER SINGLE RESPONSE
20. SUBMISSION IS: •Mandatory • Voluntary
21. REPORTING IS: • Full Coverage • Sampling
22. DATA COLLECTED IS: • Available • Restricted
23. DATA IS: • Automated • Manual
24. RESPONDENTS ARE: • Identified • Not Identified
25. DATA IS USED FOR: • Policy Making • Program Planning • Operation • Program Management and Administration • Program Evaluation • General Purpose STatistics
26. RESPONDENT CATEGORY
27. SUBJECT CATEGORIES AND SUBJECT CONTENT OF DATA COLLECTED
28. FORM OR REPORT IS: • New • Revision • Replacement • Termination

III.
Management Strategy To Design and Implement the Locator System

In view of the complexity and magnitude of the task to develop a comprehensive locator system, the substantial investments in human and financial resources, and the time required, the Commission recommends an incremental, evolutionary approach.

Moreover, the magnitude of public-use reports alone dictates the need for a pilot test. This test would encompass three or four major functional areas, such as health, education, energy, procurement, business financial information, or information on State and local governments. FEA is already underway with a locator and a dictionary, referencing data in all Federal energy programs. EPA is underway with its own directory, as are the health and education areas. Lessons learned could be used to refine the initial design before Government-wide extension.

Moving from the system concept to a cost-effective, operating system will not be easy. Nor will it be inexpensive. Important and long standing policy, social, political, and economic issues must be addressed. They include:

- The need to insure that the public understands that no personal or proprietary data will be contained within the system;
- The need to couple the use of the Directory and the Dictionary as management tools with strong congressional backing and executive leadership, including appropriate sanctions where necessary, to maximize interagency sharing once the system is put in place and begins operating;
- The need to make the system simple, both in concept and in operation; if it becomes complex, users won't use it, and data will become obsolete and unreliable;
- The need to build the system, both conceptually and operationally, carefully and incrementally; a consensus will have to be reached between conflicting interests at many milestone points that will require sensitivity and attentiveness to the unique and special needs of agencies and missions;
- The need to insure that the Dictionary, particularly, does not strangle communications. In the end, uniform terms and standard definitions can be an aid to communications, and can result in substantial cost savings and improved program effectiveness. But standardization for standardization's sake is always a real and present danger, and oftentimes leads to false economy;
- The need to solicit the involvement and participation of the many experienced and interested agencies and other organizations which have already followed this path within their own areas, both to help build a consensus by building a stake in the outcome. And to familiarize them as major users and data suppliers of the System;

It is recommended that the strategy be one of careful pilot testing in three or four selected areas with appropriate provision in the plans and timetables to solicit the views and suggestions of a broad cross-section of agencies and elements of the private sector. This incremental approach seems superior to a "grand design" because information technology itself is still rapidly evolving, and it is extremely important that involvement and commitment at agency working levels be solicited at each stage in design, development, and testing.

Building and Operating the Directory and Dictionary

The proposed file of information would be constructed in four building block stages:

- **Stage 1.** (Fact-Finding): Select the three or four pilot test areas (e.g., energy, health, education, procurement, business financial information, and information on State and local governments); investigate the current public-use reporting locator/dictionary systems and practices in use in those areas; identify key features, facilities, values, and potential problems in terms of the design of a comprehensive system.
- **Stage 2.** (Design): Design and develop a prototype locator system for pilot testing in the selected areas.
- **Stage 3.** (Evaluate Results): Evaluate the results of the pilot test, and assess changes, and effect changes, for improving the prototype system.
- **Stage 4.** (Test Pilot System): Implement and operate the pilot test system in the selected areas for a period of time, perhaps one year, to detect problems before extending the scope and use of the locator to other areas.

This approach permits periodic management assessment of both conceptual and operational problems, and permits corrections and modifications to be made before extensive resources are irrevocably committed. Additionally, it is conducive to consensus-building among participating agencies. Locator systems, data directories, and data dictionaries cannot be addressed in a vacuum, but rather must be considered in the light of their applicability within overall agency organizational missions, programs, and systems. Therefore agencies must coordinate locator specifications carefully with both their policy and operating levels to insure that the locator system's role has been properly defined within the operating environment.

IV.
Costs of the Locator

Feasibility, Development and Testing Stages (Start-up)

The estimated cost for stages 1-3, designing and testing a prototype system, would range from $1.2 to 2.7 million expended over a 42-64 month period. The range is representative of the exact number of pilot test areas selected, differences in system complexity, the degree to which existing software can be adapted, and internal coordination necessary to develop acceptable requirements and design parameters.

Operational Stages

It has been estimated that annual operating costs would average approximately $750,000 for 10 staff members and purchased services on a time-shared computer system (see Figure 6). The one year costs for pilot testing the system in selected areas (Stage 4) would approximate annual operating costs. Even though selected areas represent a small segment of reporting which the system would eventually encompass, the pilot testing would be intensive in problem solving and coordination.

The detailed cost basis for both start-up and annual operating cost estimates is presented in Figures 3 and 5. And these tasks are described in Appendix B. The cost estimates are summarized by task and function in Figures 4 and 6. The one-time start-up cost elements for stages 1, 2 and 3, represent a standard series of functions and tasks required to develop and install a system as an operational reality. It should be emphasized that these cost estimates are, at best, very broad approximations, based upon limited data. And therefore they should be regarded as only indicative of the sizeable amounts involved. The recurring annual operating costs are based on a series of tasks required to operate, maintain, and deliver services of the system. This operation is viewed strictly as a support role to agencies and to a central management organization, both of which would deal with the substance of the information contained in the system.

For comparison purposes, agencies operating similar systems were contacted to obtain costs for start-up as well as annual operation. None could extract and supply start-up costs separate from overall annual personnel costs for the organizations involved in the effort. In almost every instance, annual operating costs are embedded in staff and hardware expenses supporting a multiplicity of functions. Thus, these specific costs cannot be isolated for comparison.

FIGURE 3

One-time Start-up Costs Are Incurred For:*

- Investigation, determination, definition, and preparation of functional requirements and performance specifications (A-1);
- Investigation, assessment, and determination of available software applicability (A-2);
- Analysis, design, and preparation of detailed system design specifications (A-3);
- System implementation, including (A-4):
  - — Sub-module design,
  - — Programming,
  - — Test design and specifications,
  - — System performance testing,
  - — Acceptance testing, and
  - — System documentation.
- System installation, including (A-5):
  - — Preparation of training materials,
  - — Training,
  - — Preparation of profiles of public use reports,
  - — Data input entry for profiles,
  - — File creation,
  - — System performance assessment,
  - — System modification,
  - — System delivery.
- Determination, definition, entry, and creation of (A-6):
  - — Subject categories,
  - — Subject term vocabulary,
  - — Respondent authority list.
- Establishment of a central forms/reports reference file (A-7)

* See Appendix B for task descriptions. The references in parenthesis are to these task descriptions.

FIGURE 4

COST ESTIMATES FOR PLANNING ONE-TIME SET-UP

Function*	Duration in Months	Number of people	Labor Cost Range (000's)	Major Equipment Cost Range (000's)
Functional Requirements and Performance Specification (A-1)	12-16	8-14	280-622	
Review and Approval	1-3	8-14	23-120	
Assessment of Available Software Applicability (A-2)	6-9	8-14	140-350	
Review and Approval (A-3)	1-3	2-4	18-36	
Detailed System Design Specifications (A-4)	6-9	4-8	106-336	
Review and Approval	1-3	4-8	36-72	
System Implementation	9-12	6-12	126-325	75-100
System Installation and Training** (A-5)	6-9	10-16	90-221	150-200
Creation of Vocabularies (A-6)	42-64	2-3	81-240	30-45
Establishment of Central Forms Reference File (A-7)	3-6	1	4-8	
TOTAL	**42-64**	—	**904-2,330**	**255-345**
COMBINED LABOR AND EQUIPMENT				

* See Appendix B for Task Description.

** Labor costs for preparation of public-use profiles are borne by agencies and are not reflected in these costs. Approximately one-half hour per public-use report will yield a total of two man years of effort from all Federal agencies to prepare profiles on an estimated 10,000 reports.

FIGURE 5

Recurring Annyal Operating Costs Are Incurred For:*

- Management, administration, and agency coordination (B-1);
- Preparation of public-use reporting profiles for new or revised forms and reports (agency) (B-2);
- Input entry of public-use report profiles (agency) (B-3);
- Preparation of extract and conversion routines for entry of respondent names (agency) (B-4);
- Review, correction, and entry of modifications to agency public-use report profiles (B-5);
- Processing and storage of profiles (B-6);
- User on-line access to profiles (terminal, connect time, and system use time) (B-7);
- Servicing user requests for blank copies of forms and reports (B-8);
- Production of reports products (B-9);
- Production of one-time specialized report products (B-10);
- Revision, standardization, and maintenance of (B-11):
  - — Subject categories and subject descriptor thesaurus,
  - — Respondent category and respondent name authority list.
- Maintenance of a central forms/reports reference file (B-12);
- System modification and upgrading (B-13):
  - — Quality Assurance,
  - — Maintenance.

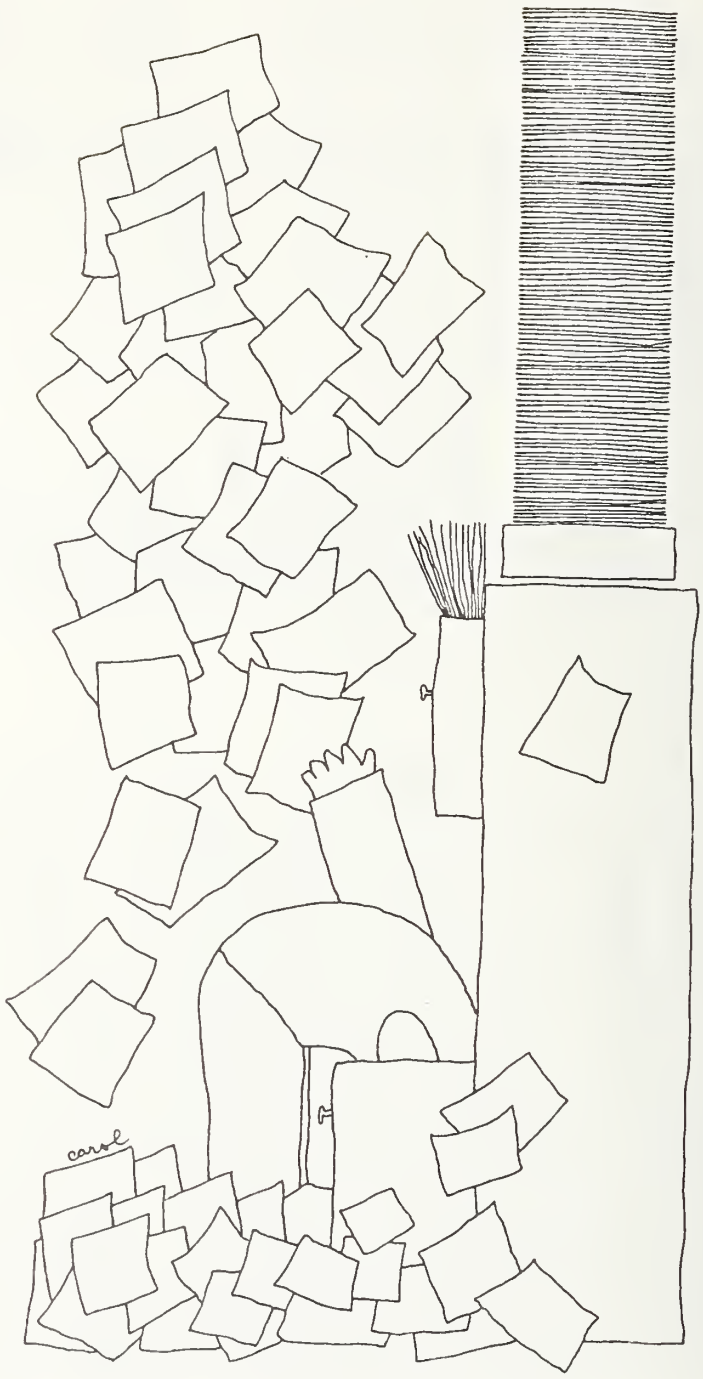
* See Appendix B for task descriptions.

FIGURE 6

COST ESTIMATES FOR PLANNING RECURRING ANNUAL OPERATION

(see Appendix B for Task Description)

Function	Staffing		Labor Cost	Major Equipment cost
	Professional	Support		
Management, Admin., and Agency Coordination (B-1)	2	1	$92,000	—
Preparation of Report Profiles (B-2)	(agency)	—	—	—
Input Entry of Profiles (B-3)	—	1	10,500	$180,000
Review and Update of Subject Descriptors (B-4)	(agency)	—	—	—
(See item 11)-Correction of Entries (B-5)	—	—	—	—
Processing and Storage of Profiles (B-6)	—	—	—	98,000
User On-Line Access (Connect and user time) (B-7)	—	1	14,500	120,000
(See item 12)-Servicing User Requests for Referenced Forms (B-8)	—	—	—	5,000
Production of Form Profile Reports (B-9)	—	—	—	36,000
(See item 13)-Production of One Time Specialized Forms (B-10)	—	—	—	40,000
Revision and Maintenance of Vocabularies (B-11)	2	—	46,000	12,000
Maintenance of Central Forms Reference File (B-12)	—	1	10,500	—
System Modification and Upgrading (B-13)	2	—	50,000	40,000
TOTAL	6	4	$22,000	$531,000
COMBINED LABOR AND EQUIPMENT-				$753,000


carol

V.
An Ultimate Locator System Concept

Once the Directory and Dictionary system are established, further opportunities would exist for tying these facilities into other management controls and management systems. Such links will provide maximum access to Federal information sources. A fully developed Federal Information Locator System would consist of three components. (See Figure 7 for a graphic illustration of system components and a description of their functions.)

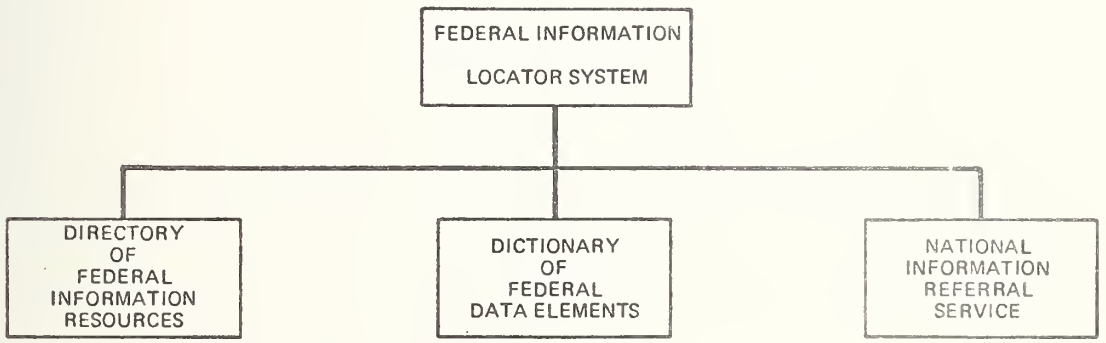
- The Directory of Federal Information Resources would be the nucleus of the system, its essential requirement. Once the Directory contains a listing of public use reporting information, it can be expanded to include:
  - — interagency reporting information;
  - — internal agency reporting information;
  - — coordination with the holdings of the States and with the private sector; and
  - — data and document holdings of Government information centers, clearinghouses, libraries, and record depositories.
- The Dictionary of Federal Data Elements would be a listing of commonly used words, terms, and names used in Federal information and reporting systems. For each entry, the dictionary would provide standard definitions where they exist as well as the idiosyncratic meanings used by specific agencies.
- The National Information Referral System would be a communications link giving people access to the locator system no matter where they may be geographically located.

The key elements in the locator system that should be immediately designed and operated are the Directory of Federal Information Resources and the Dictionary of Federal Data Elements.

The Federal Information Locator System concept and the *Congressional Sourcebook Series* have essentially the same basic objective: indentify available information resources. The *Sourcebook Series* is designed as a service to aid Congress in carrying out its oversight and budget control responsibilities. It identifies relevant and reliable information about legislated programs. The locator concept embodies this service objective as well as congressional and executive objectives of reducing Federal reporting burdens imposed on the public. The locator system, to provide these facilities, must contain profiles of more than 10,000 existing public-use reports as well as planned new reporting requirements. Identification of these resources would be extended to assist information planning and coordination at all levels of government so that public burdens arising from overlapping Federal, State and local requirements can be reduced.

FIGURE 7

THE PROPOSED FEDERAL INFORMATION LOCATOR SYSTEM


FEDERAL INFORMATION LOCATOR SYSTEM
DIRECTORY OF FEDERAL INFORMATION RESOURCES
DICTIONARY OF FEDERAL DATA ELEMENTS
NATIONAL INFORMATION REFERRAL SERVICE

A finding tool for:

- Public use information
- Inter- and intra-agency data bases
- Statistical data bases and recurrent publications
- Information holdings of government information centers and libraries
- Records and archival holdings of government
- State and local governments

Standards for:

- Data elements
- Data representations

Derived from:

- NBS
- ANSI
- Agency Data Element Directories

Terms
Names
Definitions
Abbreviations

A communication link for:

- Federal information centers and clearinghouses
- Federal libraries
- Federal archives, records centers, and repositories
- State and local libraries and records bureaus

The locator should service the public as well as the government. Information contained in *Congressional Sourcebook Series* along with other government and private sector sources should ultimately be incorporated into the locator system. The *Congressional Information Sourcebook* will thus provide an important nucleus for this service. The objective of the locator in its ultimate form is to facilitate the flow of information between the public and the government.

Scope of the Locator System

The locator, to be an effective tool for identifying duplication as well as available information resources, must encompass all reporting imposed on the public. The locator is thus viewed as a mandatory register of all Federal public use reporting. Commission studies, along with other independent studies, have revealed that a considerable number of Federal reports levied on respondents do not appear in the OMB or GAO inventories of cleared reports. Some of these reports are excluded from the clearance process while others were implemented without complying with clearance requirements. The locator must encompass these types of reports as well as those that have been cleared along with all new reporting requirements. The locator must thus be treated as a mechanism for registration and identification of reporting. The resolution of duplication and other problems arising from this reporting must be addressed as a separate issue relating to the clearance process.

If the locator is not treated as a mandatory register, the search for available data will be limited to existing cleared reports or new reports subject to clearance by OMB or GAO. The results of such a search would leave out a significant portion of Federal data collected from the public. The utility of the locator as a management, planning and coordination tool would be seriously deficient in scope and coverage.

Next Steps

The Commission recommends that a Locator System Task Force be established by the Director of OMB immediately to begin the long and difficult planning process. Initial representation is suggested in Appendix A for regular members. However, wide *ad hoc* participation in planning and system design and development is suggested because virtually every major agency in the Federal establishment has a substantial stake in the locator's design and operation. And many, beyond those listed in Appendix A, have substantial expertise in designing their own agency locators which should be tapped early, in planning stages, and later in the design, development and testing phases, including representatives of State and local governments.

Participation from the private sector is also crucial. Both individual citizens and organized sectors of the economy such as business, agriculture, and industry must be represented. To this end, a working group representing the public sector should be established as a public advisory body to the Task Force.

The Locator System Task Force should set as its first priority the design and testing of a Directory of Federal Public-use Reports in one or two of the pilot program areas (energy, health, and so on). The principal work tasks are outlined in Figure 3.

Once implemented, the locator will, of course, require routine maintenance, selective enhancement, and refinement. The key activities associated with system operation—that is, recurring cost items—are summarized in Figure 5. Initial cost estimates for these activities are given in Figure 6. And eventually, once a locator is in place and operating, many other advantages should become apparent, and their feasibility should be addressed at the appropriate time, as shown in Figure 7.

Appendix A: List Of Organizations For Potential Representation On The Directory Development Task Force (Not Inclusive)

From Government

- Office of Management and Budget
- General Accounting Office
- General Services Administration-Automated Data & Telecommunications Service (GSA/ADTS)
- General Services Administration-National Archives & Records Service (GSA/NARS)
- National Bureau of Standards, Institute for Computer Science and Technology
- Library of Congress, National Referral Center
- Library of Congress, Congressional Research Service
- National Library of Medicine
- National Technical Information Service (NTIS)
- Smithsonian Science Information Service (SSIE)
- Department of Defense, Office of the Secretary (DODOS)
- Defense Documentation Center
- Advanced Research Projects Agency (ARPA)
- Logistics Data Element Standardization and Management Office (LOGDESMO)
- Census Bureau
- National Aeronautics and Space Administration
- National Association of State Information Systems
- National Association of State Budget Officers
- National Science Foundation
- Civil Service Commission

Private Industry

- Information Industry Association (IIA)
- American Federation of Information Processing Societies (AFIPS)
- American Society for Information Science (ASIS)
- American National Standards Institute (ANSI)
- Business Equipment Manufacturers Association (BEMA)
- Association of Records Manuals and Administrations (ARMA)

Public Representatives

- Consumers
- Privacy experts
- Other groups

Appendix B: Task Descriptions

One-time Start-up Costs

Functional Requirements and Performance Specifications (A-1)

This set of specifications, in the form of a feasibility study, will be the primary operational product of the development task force. It will describe what the system is to consist of, and what it must do as well as not do. It will be the basis for subsequent software assessment and system design. It would represent the combined views of agency participants, the task force chairman, and the final approval authorities. The costs represent a full-time task force professional staff, with clerical support, ranging from eight to fourteen people for a duration of 12 to 16 months. Time and costs for participating agency membership on the task force and parent agency review are not included.

Assessment of Software Applicability (A-2)

This function includes an assessment of currently available, operational software systems that can meet the requirements and performance specifications for the directory. The assessment would include issuance of an RFP containing specifications and evaluation of proposals submitted. Solicitation of proposals would include selection of a software system, as well as an operating hardware environment. The latter is assumed to be obtainable from among existing time-sharing systems owned, leased, or contracted for on a service basis by the Federal Government.

The assessment, including preparation and issuance of an RFP, evaluation of results, and selection would be performed by the task force within a 6 to 9 month period.

Detailed System Design Specifications (A-3)

Based on functional requirements and performance specifications, and the selected software system and operating hardware environment, the system will be designed and detailed specifications prepared.

This function may be performed by the task force, by personnel selected from agency resources, or by assistance of a contractor. For the latter two cases, it is assumed that a component of the task force will oversee the design activity and results. The system will be implemented using the approved design specifications.

System Implementation (A-4)

This function may be performed by either Federal personnel resources or contractor personnel or some combination of both. It will be managed by the task force. The basic assumption made for this function is that an existing software system or package, with minimal modification or definition for directory application, will be used. A major portion of the implementation may encompass system parameter definition of recurring as well as ad-hoc demand reporting capability. File maintenance, index creation and keying capabilities will be existing features of the software selected.

Implementation will be based on test data as well as a sampling of actual cases. The test material should be developed by the task force and made available during submodule design.

System Installation (A-5)

This function will overlap with the preceding implementation task and will be performed in parallel with the subsequent task of vocabulary creation for subject terms and respondents. This function should be performed with a portion of the group used for system implementation.

The most intensive and time consuming aspect will be the preparation of public use reporting profiles, entry, creation of a data base, and review and correction of the results. It is assumed that responsible agencies will prepare their own report profiles based on guidelines and materials furnished by the task force during training sessions. The review, data entry, and correction of these profiles will be conducted centrally. Results of corrections and modifications will be returned to agencies to provide examples for improvement in quality of subsequent profiles. Training costs for presence of agency personnel at training sessions are not included. Also not included are the agency costs for preparation of their form profiles. A range of 1 to 2 hours per public use report profile can be used. For this estimate as well as data entry, an upward limit of 10,000 public use reports is assumed. This would include those reports under the responsibility of OMB and GAO, those excluded such as IRS, and those which may be bypassing present clearance procedures. An actual form or report count estimate is used in lieu of clearance cases at OMB since many single cases contain multiple forms. It is also assumed that data from profiles will be entered in free form as off-line batched input. Corrections will be made using on-line, interactive facilities of the system.

After the "live" data base has been created from actual profiles, the system would be operated on a turn-key basis by the group responsible for installation. During this period, the performance will be monitored by the task force, change modifications will be defined and implemented, and the system will be accepted for delivery as operational.

Creation of Vocabulary and Authority Lists (A-6)

The foundation for this function will have been developed during performance of the requirements function and refined during the design function. This foundation consists of developing a set of subject categories and subject terms or descriptors based on actual examination of public-use reports. The "tentative" vocabulary would be issued to agencies for preparation of report profiles. After these profiles have been entered into the directory system along with terms or descriptors suggested by the agency, they would be reviewed and modified to prepare a thesaurus of authoritative public use reporting descriptors.

This function must be performed by a group experienced in subject indexing and thesaurus building. The skills are identical to those found, for example, in the Defense Documentation Center, National Technical Information Service, NASA's Scientific and Technical Information Facility, the Library of Congress, or commercial abstracting and indexing services.

Initially, during the requirements and design stages, the work of this group should be based on suggestions from agency program and subjecting indexing personnel where available. The actual review of public use forms by the central group to develop subject terms will consist principally of a review of material prepared and submitted by agencies. For ease of focusing on the efforts and estimated costs, all preparatory work supporting the vocabulary aspects of the requirements and design functions are included under this single function. It is assumed that this group of specialists will be available for the duration of the start-up from requirements determination through final installation, delivery, and operation of the system. A nucleus of this group would be retained for recurring annual operation to perform vocabulary control and review of agency profile inputs. The respondent categories and classification schemes will also be handled by this group.

It is assumed that the system will have a global correction capability for both subject terms and respondent categories. Thus inconsistencies creating stray entries in a composite index could be corrected in the original source profiles that generate these entries.

Not included in these estimates is agency assistance in developing subject terms and classifications that describe the types of respondents to reports.

Considering the billions upon billions of dollars the Federal Government has invested in its data and literature holdings, expending a figure of $1—3 million to develop a tool to help officials and citizens find their way through the labyrinth of filing cabinets, computer data banks, library bookshelves and dusty archives does not seem, to the Commission at least, to be an inordinately high number. Nor does three to five years development time seem excessive, given the enormity of the task and the need to carefully build a consensus at key decision points.

Establishment of a Central Form/Reports Reference File (A-7)

This one-time function involves setting up a file in public use report serial number sequence. Each file will include a copy of the report or form, the profile sheet submitted by the agency, and copies of correspondence and related materials residing in present OMB and GAO files.

Recurring Annual Operating Costs

Cost estimates in Figure 6 are based on annual professional and non-professional staffing level and major equipment costs for performance of each function. The workload level includes an assumption of 1,000 new, planned report requirements per year. The number of requests for services from directory facilities have not been made. The service functions have been identified with a minimum staffing level indicated. The performance of some functions are similar one to another. In these instances the staffing and costs have been shown for a single function with an appropriate cross-reference from the other function.

Basic assumptions for annual operation and staffing are as follows:

Management, Administration and Agency Coordination (B-1)

This function includes management of directory facilities and operations, coordinating public use reporting profile requirements with OMB and agency management, coordinating the entry and maintenance of reporting profiles with agency program planning and information management, managing vocabulary control for indexing and retrieval, managing system maintenance, and planning for inclusion of other information resources in the directory.

Preparation of Public Use Report Profiles (B-2)

This function is performed by agencies when planning new report requirements. It consists of completing the Public Use Report Profile. Costs are not estimated, although 1—2 hours each for 1,000 new forms can be used as an estimate of labor requirements.

Input Entry of Public Use Report Profiles (B-3)

Where agency volume warrants a terminal, profiles would be entered directly by the agencies. Labor costs have not been included although a half hour each for 1,000 new reports can be used as an outside labor estimate. Provision has been made for a single data entry person at the central location to process input for those agencies without terminals. For equipment costs, rental of 50 terminals is included with connect time and file access costs.

Review and Update of Subject Descriptors and Respondent Categories (B-4)

This function is performed by agency personnel working in conjunction with the locator vocabulary specialists. The objective of this function is to improve the quality of indexing and thus location of information based on results from inquiries to the system. New terms would be added to the vocabulary as well as terms that more specifically describe subject content.

Adjustment of the vocabulary will optimize search results by encompassing all reports relevant to an inquiry and minimize the number of non-relevant identifications. Adjustments to respondent categories will be used to produce better profiles of reporting burdens on segments of the public. Agency participation in this fine-tuning process is essential to effectiveness of the locator as a finding aid.

Review, Correction, and Entry of Modifications to Agency Public Use Report Profiles (B-5)

This function is performed by the directory staff after profiles have been entered into the system. Staffing and cost estimates for this function are included in "Revision, Standardization, and Maintenance" (item B-11).

Processing and Storage of Profiles (B-6)

This function involves only hardware costs. It includes CPU time and annual file storage costs. Principally the function consists of reorganizing files and creating inverted indexes for direct access to any item profiling forms.

User On-Line Access to Profiles (B-7)

This function is performed by agency users where terminals are available. Provision is made for directory staff support where terminals are not available in agencies for OMB information management offices. Total hardware costs for access are estimated. These costs include listing results from an inquiry.

Servicing User Requests for Blank Copies of Forms and Reports (B-8)

Initially the directory services will include furnishing copies of forms referenced by an inquiry. Hard copies of these forms will be maintained in a central reference file until such time that microfiche sets become feasible for distribution to large volume inquiries.

Hardware costs for forms reproduction service are shown. However, labor costs for this service are shown in "Maintenance of a Central Forms/Reports Reference File" (item B-12).

Production of Report Products (B-9)

This function involves periodic production of standard recurring report products. It is essentially a self-generating function that involves no significant labor costs. Labor support, when required, will be covered by the staffing for function "System Modification and Upgrading" (item B-13). The hardware costs do not include reproduction for distribution outside OMB and each agency information management office.

Production of One-Time Specialized Report Products (B-10)

These reports are essentially those that have not been anticipated in advance and must therefore be developed to answer specific requests. Staffing to support design, development and generation of these reports is included in function "System Modification and Upgrading" (item B-13).

The costs shown for hardware include utilization to develop and produce these one-time reports.

Revision, Standardization, and Maintenance of Subject and Respondent Descriptors (B-11)

This function is performed by experienced indexing, vocabulary control, and thesaurus development personnel. It consists of maintaining consistency in both subject terms or descriptors and respondent categories to prevent spurious references. It also includes the development of new categories and classifications as well as insertion of "blind" or "transparent" cross-references in the dictionary. Hardware costs are included for access to modify and maintain the vocabulary. This function will diminish in need after the first three years of operation.

Maintenance of a central Forms/Reports Reference File (B-12)

This function involves control, maintenance, and services from the hard-copy reference file of public use forms. These files include a copy of the blank form, the form profile, agency correspondence coordinating the need and exchange of data with other agencies, coordination correspondence, containing respondent comments, an impact statement, and justifications for not using existing data resources or incorporating respondent recommendations into the new reporting requirement.

System Modification and Upgrading (B-13)

This function will be performed by personnel experienced in use of the software system selected for the directory. It includes trouble shooting modifications to the system to correct problems, upgrade capabilities, or incorporation of changes requested by management. The staffing will also support development and generation of one-time specialized reports.

Appendix C: Exemplary Locator Systems Examined

EXEMPLARY LOCATOR SYSTEMS*	PROPONENT ORGANIZATIONS*
DATA CATALOGUE	SYNERGETICS CORPORATION
DATA DICTIONARY	CINCOM SYSTEMS, INC
DATA DICTIONARY/DIRECTORY	IBM
DATAMANAGER	MSP, INC
DB/DC DATA DICTIONARY SYSTEM	IBM
LEXICON	ARTHUR ANDERSEN AND COMPANY
UCC TEN DATA DICTIONARY	UNIVERSITY COMPUTING COMPANY
DARCOM DATA ELEMENT DICTIONARY	DEPARTMENT OF DEFENSE - ARMY
INDEX TO CIVIL SERVICE COMMISSION PUBLICATIONS	CIVIL SERVICE COMMISSION
CIVIL AERONAUTICS BOARD DATA INDEX	CIVIL AERONAUTICS BOARD
D-20 DATA ELEMENT DICTIONARY	DEPARTMENT OF TRANSPORTATION
INFORMATION ROMTS CONTROL AUTOMATED SYSTEM	DEPARTMENT OF DEFENSE - ARMY
FEDERAL ENERGY LOCATOR SYSTEM	FEDERAL ENERGY ADMINISTRATION**
GENERAL INFORMATION MANAGEMENT SYSTEM	COMPUTER SCIENCE CORPORATION
ENVIRONMENTAL INFORMATION SYSTEMS DIRECTORY	ENVIRONMENTAL PROTECTION AGENCY
RECORDS ASSOCIATION SYSTEM II	DEPARTMENT OF DEFENSE - NAVY
DEFENSE INTEGRATED DATA SYSTEM	DEPARTMENT OF DEFENSE
NATIONAL REFERRAL CENTER DIRECTORY	NATIONAL REFERRAL CENTER
AIDS DATA ELEMENT MANAGEMENT SYSTEM	AGENCY FOR INTERNATIONAL DEVELOPMENT
VETERANS ADMINISTRATION DATA ELEMENT DICTIONARY	VETERANS ADMINISTRATION
DATA LABLE CONTROL SYSTEM	US COAST GUARD
CATALOGING AND INDEXING SYSTEM	DEPARTMENT OF AGRICULTURE
CURRENT RESEARCH INFORMATION SYSTEM	DEPARTMENT OF AGRICULTURE
NATIONAL TECHNICAL INFORMATION SERVICE SYSTEM	DEPARTMENT OF COMMERCE
DEFENSE DOCUMENTATION CENTER SYSTEM	DEPARTMENT OF DEFENSE
ARMY RESEARCH AND DEVELOPMENT INFORMATION SYSTEM	DEPARTMENT OF DEFENSE - ARMY
MEDICAL LITERATURE ANALYSIS AND RETRIEVAL SYSTEMS	DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
WATER DATA SYSTEM	DEPARTMENT OF INTERIOR
WATER RESOURCES SCIENTIFIC INFORMATION CENTER SYSTEM	DEPARTMENT OF INTERIOR
HIGHWAY RESEARCH INFORMATION SYSTEM	DEPARTMENT OF TRANSPORTATION
TRANSPORTATION NOISE RESEARCH INFORMATION SERVICE	DEPARTMENT OF TRANSPORTATION
TECHNICAL INFORMATION EXTENSION SYSTEM	ATOMIC ENERGY COMMISSION**
NUCLEAR DESALINATION INFORMATION CENTER SYSTEM	ATOMIC ENERGY COMMISSION**
SCIENTIFIC AND TECHNICAL INFORMATION SYSTEM	NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
NATURAL HISTORY INFORMATION SYSTEM	SMITHSONIAN INSTITUTION
NATIONAL SCIENCE INFORMATION EXCHANGE	SMITHSONIAN INSTITUTION
LOGISTICS DATA RESOURCE MANAGEMENT SYSTEM	DEPARTMENT OF DEFENSE

* Some system titles and organization names are abbreviated.
** Subsumed in Department of Energy.

THE URGENT NEED FOR ENERGY INFORMATION STANDARDS (TO "Btu" OR NOT TO "Btu")

Richard W. Kline, Jr

Office of Energy Systems Data
Energy Information and Analysis
Federal Energy Administration
Washington, D. C. 20461

1. Background

In 1973, the Arab oil embargo forced the United States, and the world, to look closely at the problems of obtaining oil in the face of scarcity. This in turn precipitated national concern over the total energy picture here in the United States. The spectrum of energy related problems -- supply, resources, demand, distribution, conservation, price structures, etc. -- received unprecedented attention and the development of policies and programs directed toward providing near and longterm solutions assumed high priority. Of course, one of the first tasks involved gathering information about energy.

The need for large amounts of energy information was widespread and the Federal government became, probably, the most active sector in collecting data about energy. Jurisdictional responsibilities associated with the energy crisis were spread throughout Federal departments, agencies, administrations, and commissions. What resulted from this decentralized management of the nation's energy problem was, among other things, a proliferation of energy information gathering activities. There was, for the most part, a lack of coordination both within the Federal government and between the Federal government and the energy industry. Wide variances in reported data cropped up everywhere. As an example, in January 1975 official estimates of American imports of oil ranged from 6.3 million barrels per day to 9.1 million barrels per day. That's a difference of over 50% in less than four weeks time and was hardly the ideal situation for developing policy and programs which deal with U.S. dependence on imports of Arab oil. The credibility of energy statistics published by the Federal government became suspect and it followed that the policies and programs which relied heavily on a credible information base became equally suspect. The public, the energy industry, and the Congress openly criticized this lack of credibility - and rightly so.

Senator Jackson of Washington summed it up when he stated* during hearings on energy problems that:

> "We still do not have the facts we need to make sound national economic and energy policy....We don't have accurate figures on stocks, on demand, on costs, on imports, or virtually anything else ....most of the confusion surrounding energy policy ...is the direct result of inaccurate or inadequate information."

It was painfully obvious that effective management of the energy crisis would be closely tied to the development of a reliable, consistent, and credible information base.

One of the first positive measures taken toward improving the quality of our energy information was the formation of the Federal Interagency Council on Energy Information (FICEI) in December of 1975. The FICEI serves to identify and initiate programs for coordinating and improving the energy information activities within the Federal government. Later, in August of 1976, Congress passed the Energy Conservation and Production Act (ECPA). The ECPA has, among its provisions, a specific directive which created the Office of Energy Information and Analysis (EI&A) within the Federal Energy Administration. EI&A was charged with developing a National Energy Information System to meet the nation's requirements for reliable, coordinated, and credible energy information. President Carter's National Energy Plan (April 1977) specifically addresses the need for "reliable information on energy matters." And most recently, the Department of Energy Organization Act, passed in August 1977, further emphasizes the role of and requirement for high quality energy information. In short, the Federal government has recognized the problem and is setting up the machinery to analyze and solve it.

2. The Problems

The problem is not that we do not collect enough energy information. Preliminary survey results show that the Federal government has over 250 energy information systems in operation. So, quantity is not the problem. A great deal of data is transferred from the energy industry to the Federal government and within the Federal government, but the communication of the meaning -- the understanding -- is often not. Problem number one, then, is the clarification of misunderstood or ambiguous information.

As an example, let's consider the reporting of "inventories of residual fuel oil" for a particular month. On the surface, this seems to be a fairly self explanatory piece of information and reporting this data would seem to be relatively straight forward -- not so, unfortunately. Some definitions for residual fuel oil include only No. 5 and No. 6 grade fuel oil as residual fuel oil while others include No. 4 in addition to No. 5 and No. 6 grade fuel oil. Often, no definition at all is provided. But, if we finally get the definitional problems ironed out, can we then be confident that we know what the inventories of residual fuel oil are? Not yet. Some firms report as "inventory" those products which they have contracted to buy but, possibly, have not physically received as of the day the report is rendered. Given that we can also clarify the intended meaning of the term "inventory", we may still not have the complete understanding of the volume data reported.

*U.S. Congress, Senate, Committee on Government Operations Current Energy Shortages Oversight Series, The Federal Energy Office: Hearings Before the Permanent Subcommittee on Investigations, 93D Congress, 1st Sess., 25 January 1974. Part 5. Page 597.

Since the physical characteristics of petroleum products change with temperature, we would also need to know if the reported volume data has been normalized to a reference temperature and pressure. Failure to account for pressure and temperature variations can induce errors up to 10% or more. There are many more stumbling blocks which must be overcome if we hope to gain a full understanding of the data. The point to be made here is that information (data) may be exchanged but, through a lack of specificity, the exact meaning of the information is not exchanged. This places the burden of "filling in the blanks" upon the recipient of the data and opens the door for introducing large errors.

A second problem with our energy information involves duplication -- or industry and government needlessly. In the case of "near-duplications," the problems of data comparability are compounded. Often we discover data collection systems which appear to duplicate one another, but when we cross-check one against the other we find they are not in close agreement. This weakens our confidence in the data of both systems and further tarnishes the credibility of Federal programs based upon the data. Actually, each system may be producing quality data, but, because we do not know the specifications which drive the system's performance, comparisons between the output data fall apart.

A third problem centers around the triple role which the Federal government assumes when it (1) performs its energy regulatory function, (2) develops policies and plans to manage national energy programs, and (3) conducts research and development aimed at solutions to energy problems. The types of information collected to support these three functions differ in their level of detail and focus. These differences are often not recognized and frequently result in errors due to improper use of the information.

Data systems designed for regulatory purposes characteristically focus upon a very narrow band of information; for example, "increased crude oil costs for Aviation Jet Fuel incurred two or more months before the reporting period and not recovered through the reporting period". Statistical systems address broader questions such as average costs for all crude oil, while research and development oriented data systems may seek information about crude oil production costs using advanced high technology recovery systems. Each system is getting data about "crude oil costs", but that is where the similarity ends. Accordingly the energy information community must find ways to keep these data distinct and thereby discourage improper use.

There are, no doubt, other problems which should command as much attention as these I have listed (we seem to have no problem with the supply of energy problems -- just the energy), but these three specific problems share a commonality in that they can be solved, for the most part, through the application of an effective energy information standards program.

3. The Solutions

In order to establish a credible base of energy information it will be necessary to eliminate definitional ambiguities, consolidate the duplications, and increase the specificity with which we collect and use energy information. A comprehensive energy information standards program seems to offer the most promise for getting "on-top" of these problems early. Standards provide the rigid detail required to preserve clarity and precision in the terminology and definitions, establish a basis for comparing and controlling the independent generation of new and potentially duplicative energy information systems, and mandate the specificity required to guide our collection and publication efforts away from improper use of the energy information.

More specifically, what is called for is a program which includes standards for: terminology, nomenclature, classification and coding structures, data elements, data items, units of measure, conversion factors, analytical methodologies, data collection/publication formats, and disclosure of data sources and data quality. Such a program, though perhaps not exhaustive, will provide the set of "tools" necessary to work on the energy information credibility problem.

Standards for terminology and definitions will remove the ambiguities associated with collecting and reporting energy information. Providing specific, comprehensive, and accepted definitions for the basic vocabulary of energy information establishes the foundation for efficient and effective communication. Once this basic vocabulary is standardized, the building blocks are available to standardize the data elements and data items which are used in the data systems. These are, of course, the most critical parts of the standardization puzzle because they represent the energy information questions that are asked and the responses that are supplied. At this level, communication must be precise. Otherwise, the objective of the standards program is lost.

Units of measure, conversion factors, analytical methodologies, and operational measurement procedures cannot be overlooked. After all, most of the energy information collected is numerical data. Imprecise measurements, analyses, and conversions will completely negate the large effort expended to gain a precise definitional base. As an example, much of the value of knowing precisely what is meant by "U.S. imports of Arab oil" is lost if we have no precise means to determine "how much" there is.

A frequently overlooked need is the requirement to have standards covering disclosure of data sources and data quality. In other words, "Where did you get the data?" and "How good is it?" Often the user gives only passing notice to either question until his program "slows up" because of bad data. Then the questions of "Where from?" and "How good?" are suddenly of great concern.

In the name of humane treatment for both the respondents to energy information questionnaires and the end users of energy information publications, I have added the requirement for data collection and publication formats. Everyone seems to be collecting energy information these days, and, if they do collect it, it's a good bet that they will publish it in at least some unofficial regard. Some common format for collecting and publishing energy statistics would certainly be welcomed. The goal, of course, is to make data easier to supply, collect, and use for all segments of the energy information community.

So much for what we need by way of energy information standards. The next question is how do we get them developed and working?

A review of the strengths and weaknesses of other standards programs leads to the following set of conclusions:

a. Standards must be produced in an enviroment which encourages the participation of, and comment by, all segments of the community impacted by the standard.

b. Standards must be produced in a timely fashion in order to prevent their becoming obsolete before they become operational. As Federal energy information requirements change, energy information standards must provide a current and stable foundation to ensure high quality information systems.

c. Standards must be produced through a centralized coordinating mechanism with sufficient influence to resolve conflicts and maintain the standards program on schedule.

d. Standards must be developed which are capable of operating in the automatic data processing (ADP) environment.

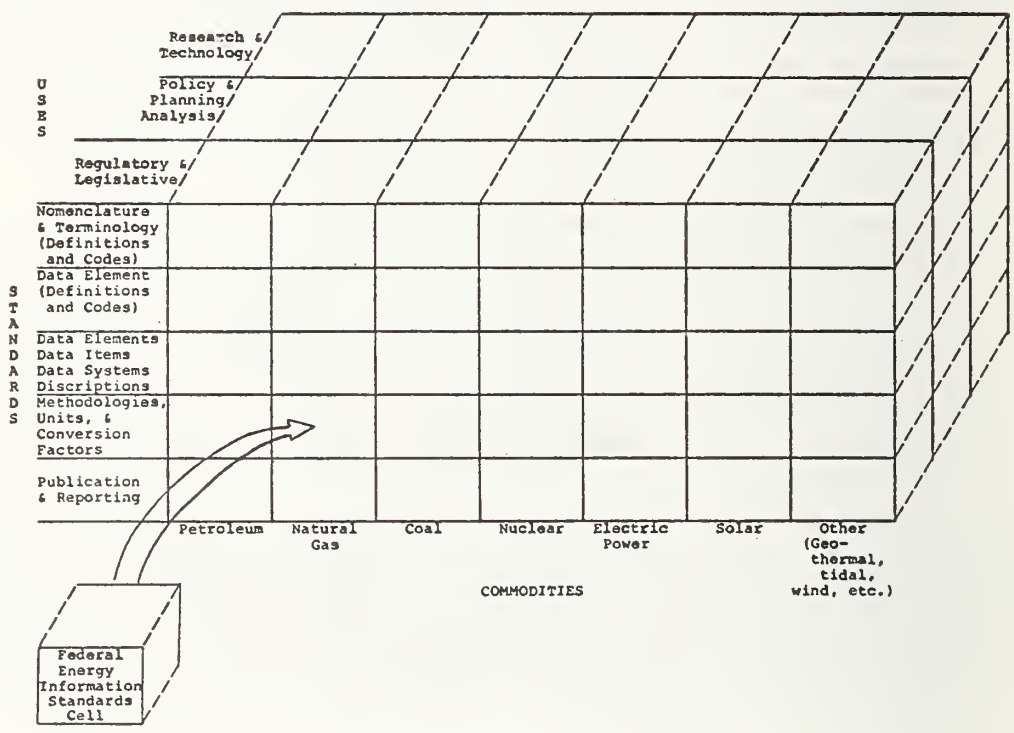
e. Standards must be developed to provide sufficiently comprehensive coverage of the subject area without becoming "diluted" in an attempt to "stretch" applicability beyond the necessary bounds.

f. Standards, once developed, must be maintained and updated as the requirements for information shift.

Recently, the Federal Interagency Council on Energy Information put together a plan for the development of a comprehensive set of energy information standards such as we have discussed here. The approach breaks the overall energy information problem into three dimensions: energy information uses, energy information commodities, and energy information standards. Using these three dimensions, the program subdivides into energy information standards cells as shown in Figure 1. In this fashion, a large undertaking is broken into smaller efforts, allowing for multiple coordination within and across the three dimensions of a structured problem.

Of course, once the candidate organizational framework is developed, the next question is "Who will do the job?" The Department of Energy seems to be the logical choice and, as a matter of fact, the ground work is already being laid for this task. But, the Department of Energy can not do it alone. Experience has shown that standards developed unilaterally, and in isolation, are by and large ignored. As such, the participation of the industry, the ADP community, the state and local governments, as well as the Federal government is required. DOE is in the best position to take the lead and manage this program, make the critical decisions necessary to keep it going and productive, and provide the resources to see it through to the end. However, the standards generation activities must be a team effort on the part of the entire energy information community. The firsthand expertise of the energy industry, a working knowledge of the energy information requirements of government, and the experience of the data processing community must be tapped and coordinated if any success is to be achieved. Standards development is not a limelight activity and standards work is by and large a thankless task. Nevertheless, the energy information community is now acutely aware of the urgent need for energy information standards and the time to get started is now.

Figure 1. Federal Interagency Council on Energy Information Standards Cell Matrix



Criteria for the Selection of Data Dictionary/Directory Systems

Bernard K. Plagman

DBD Systems, Inc.
77 No. Centre Avenue
Rockville Centre, NY 11570
(516) 678-1447

With the recognition that data is a resource with corporate-wide impact and policy implications, many organizations are considering various tools to assist in the task of administration and control of data. The most important of these tools is the Data Dictionary/Directory System. Many commercial software products are available to fulfill this need with varying characteristics and capabilities. This paper presents an approach to the selection and evaluation of a DD/DS including a uniform set of evaluation criteria for the comparison of these packaged software systems.

1. Introduction

A Data Dictionary/Directory System (DD/DS) is a software package designed to assist the Data Base Administration function of an organization in the coordination and control of data resources. It has been formerly defined as;[1]

"The central repository of information about data in an organization."

The DD/DS can be viewed as an application whose primary users are systems and end-user personnel involved in the shared use of data resources. The advent of data base management systems (DBMS) has allowed users of data to share a single physical representation of a data base without undue sacrifice in operating efficiencies. The use of the DBMS has made this possible at a technical level by delivering data independence in the form of separation of the physical representation of data from the logical user's view of the same physical representation. However, the DBMS alone cannot coordinate and control the shared use of data resources. To accomplish this it is necessary to support data resources in a manner similar to the way other corporate resources are managed.

Typical resources found in most organizations include money and merchandise. Management in instituting programs to effectively utilize these resources find it necessary to appoint individuals to be responsible for the coordination and control of each respective resource. Thus, a Controller is responsible for the organization's financial resources and the Merchandising Manager for its inventory. These individuals, in turn, have found it difficult, if not impossible, to fulfill their jobs without the support of automated systems to help manage the resource. Thus the emergence of Financial Control Systems and Inventory Management Systems.

---

[1] Plagman, B., Data Dictionary/Directory System: A Tool for Administration and Control. Auerbach Data Base Management Series, #22-01-02, 1977.

The recognition in recent years that an organization ought to have a Data Base Administration (DBA) function responsible for the coordination and control of data related activities in the enterprise, was borne out of the cognizance of data as a corporate resource. The realization, soon thereafter, that the DBA function could not effectively operate without the automated support of a data inventory system, (i.e., a data dictionary/directory system) is a logical progression of events. For many years, (from 1970 - 1974) commercial DD/DS's being generally unavailable, most users found it necessary to improvise and develop "home-grown" systems. Such an effort would typically take nine to twelve (9-12) months and cost from $40,000 - $100,000, depending on the sophistication of the effort. Today, a survey of the market place (see figure 1) uncovers better than half a dozen DD/DS packages of one form or another. These packages range in cost from $10,000 - $25,000. Applying the logic of "MAKE VS BUY" analysis (see figure 2), most organizations considering a DD/DS are shopping and comparing the available commercial offerings.

This paper addresses the problem of how to compare, evaluate and select a DD/DS from amongst the available commercial offerings, to insure a "best fit" in a particular data base environment.

The recommended approach for DD/DS package selection is to establish a selection methodology which will minimize the subjectivity and bias of those involved in the selection process. The steps in this methodology are outlined in figure 3 and expanded in greater detail in the remaining sections of this portfolio.

2. DD/DS Selection Methodology

The first steps in evaluating and selecting a DD/DS package are oriented towards planning the effort. Attention is directed towards enumerating the steps to be followed in the selection methodology (see figure 3) and estimating time required for the completion of tasks. Selection of personnel to perform these tasks should be made on the basis of each person's contribution to the selection effort. A typical selection team might include five persons representing the following areas within the organization.

Data Base Administration (Chairperson)
Technical Services
Applications Programming
Systems Design
Operations

With wide representation the selection team can be expected to avoid the bias of any one particular viewpoint. For example, were Technical Services choosing a DD/DS, the choice might be a system heavily oriented towards the technical support of the installation's DBMS, with lesser regard for the data definition capabilities of the package. This equilibrium of viewpoints is important to establish and is accomplished by using the weighting mechanism described below.

In conjunction with establishing the selection methodology, its participants must initiate an education and training program to accomplish two objectives:

(1) Make decision-making management within the EDP area aware of DD/DS trade-offs involved in the selection, and

(2) Provide the selection team with enough information to procede with the evaluation.

The source of this education will be, for the most part, in the literature. References are provided at the end of this paper. In addition, we recommend highly seminars offered by independent vendors.

3. DD/DS Selection Criteria

The evaluation and selection of a DD/DS package is based upon the comparison of relevant packages based upon a uniform set of evaluation criteria. The specification of this uniform set of criteria is an important step in the selection process.

The uniform DE/DS selection criteria is comprised of eight (8) major selection criteria categories (see figure 4). Each of these is segmented further to adequately describe DD/DS capabilities.

[1.0] Data Description Facilities

- [1.1] Type of Input (Keyword versus Preformatted)
- [1.2] Entity Structure Definition Support
- [1.3] Mandatory/Optional Definitions
- [1.4] Default Functions
- [1.5] Reference Definition
- [1.6] Generation of DD/DS Input from Host Language Code

[2.0] Data Documentation Support

- [2.1] Identification Attributes
- [2.2] Source Attributes
- [2.3] Type of Data Attributes (Generic classifications)
- [2.4] Usage/User Attributes
- [2.5] Qualification Attributes
- [2.6] Relationship Attributes
- [2.7] Attributes Describing System Entities
- [2.8] Other Documentation Considerations (e.g., Keyword capability, version control, etc.)

[3.0] Data Description Generation

- [3.1] Generation of Host Language Data Description
- [3.2] Generation of DBMS DDL
- [3.3] Interface Architecture (i.e., the design of interfaces)
- [3.4] Completeness of Interfaces

[4.0] Security Support

- [4.1] Attributes for Describing User Data Base Security
- [4.2] Security Support for the DD/DS Data Base

[5.0] Integrity Support

- [5.1] Edit and Validation
- [5.2] Audit/Derivation Functions
- [5.3] Test Data Generation
- [5.4] Distributed Data Base Support

[6.0] User Interfaces/Outputs

- [6.1] Standard Reports
- [6.2] Ad Hoc Query/Update
- [6.3] On-Line Facilities (Retrieval/Update)

[7.0] Ease of Use and Resource Utilization

- [7.1] Ease of Learning
- [7.2] Level of Expertise Required
- [7.3] Compatability with Existing Procedures
- [7.4] Operational Impact

[8.0] Vendor Support

- [8.1] Vendor Stability
- [8.2] Commitment to Package
- [8.3] Reliability and Quality of Support
- [8.4] User's Group
- [8.5] Documentation

3.1 Mandatory Criteria

In identifying and describing the uniform selection criteria according to the foregoing guidelines, it is necessary to identify specific criteria which will be treated as mandatory. DD/DS packages which fail to meet minimum requirements for mandatory criteria would be eliminated from the evaluation process at this early stage. For example; under the criteria [1.6] of Data Description Facilities, Generation of DD/DS Input from Host Language Code, it may be deemed necessary to have a COBOL program scanner which would generate input for the DD/DS. This type of facility may be mandatory due to anticipated major conversion activity. Thus DD/DS packages without this facility would be eliminated from contention (e.g., LEXICON).

An integral part of the elimination of alternatives based upon mandatory criteria is the classification of available DD/DS packages based upon their dependency on a DBMS and/or the relative degree of control afforded to the DBA upon integration into the data base environment. A given DD/DS package must utilize some mechanism to manage, i.e., organize, access and control, the DD/DS data base. This meta data base is often times managed by a commercial DBMS, thus requiring the installation of the DBMS in order to run the DD/DS package. This can, of course, be an advantage if the meta data base and the DBMS directories are actually one and the same. Additionally, in situations where the organization is absolutely committed to the underlying DBMS, the fact that the DD/DS is dependent on the DBMS would be an advantage.

Nevertheless, many DD/DS packages remain independent of any particular DBMS, allowing users to be somewhat more flexible with regard to a DBMS.

Regardless of whether or not a DD/DS is dependent or independent of a DBMS, it can still provide a high degree of active control over the data base environment. This is accomplished by providing automated interfaces for the generation of data descriptions (see criteria 3.0 above). In doing so, the specification of all data descriptions can be controlled through the DD/DS. Packages not providing sufficient degrees of active control are referred to as passive DD/DS's. An alternative name for the active controlling DD/DS, is the "integrated" DD/DS.

Figure 5 illustrates a typical classification of DD/DS packages in terms of:

Dependence versus independence
Relative active control versus passiveness.

It should be noted that in figure 5 we have assumed that a DD/DS package is active (integrated), only if it includes a facility for generating DDL for at least one DBMS. Also, we have assumed that if a DD/DS package requires a DBMS's access methods to operate, it is classified as dependent, even if the vendor doesn't require the prior purchase of the DBMS in question.

4. Weighting of the Selection Criteria

Having arrived at a set of uniform DD/DS selection criteria, the next important step is to assign weights to each of the criteria. This process represents the "tailoring" of the criteria to the particular needs of the data base environment in question.

The mechanical process of assigning weights is similar to the calculation of a weighted average and is hypothetically illustrated in Figure 6. Notice that weights are applied successively at each level of the criteria.

The task of deciding upon the values for the weights is difficult and understandably critical. This task is the single most important activity of the DD/DS selection team. Experience has shown that a workable approach is to have each team member assign a full set of weights individually and then convene as a group to negotiate differences.

The process of applying weights will thus reflect the need for particular features of a DD/DS. For example; an installation may already be using one or more DBMS's or a program library manager (e.g., Librarian) and special interfaces would be advantageous. The weights in Category Three (3) should reflect these requirements. Another installation may be using TSO for on-line programming and therefore require a DD/DS with compatable on-line capability. The weights in Category Six (6) should reflect this situation.

5. Evaluation of DD/DS Packages

At this stage of the evaluation and selection process a uniform set of criteria has been established and based upon mandatory criteria a specific group of DD/DS packages have been designated for detailed analysis. The next level of analysis is for the most part technical and is completely independent of the particular needs of organization. At this point the evaluation concentrates on scoring each package against the uniform selection criteria. Ideally, this task need not be performed by the selection team. By assigning different individuals to do this scoring an element of objectivity is introduced into the selection.

There are various sources of information available to develop these scores. Vendors, of course, are always ready, willing and able to participate, however one must be extremely careful of ambiguities and errors of omission. Independent sources of information are available in the literature (see references) or from consulting firms actively involved in the evaluation and implementation of DD/DS packages. An advantage to using external services, in this regard, is that time is not wasted learning packages which will not be used.

6. Final Selection of a DD/DS

The final step in the evaluation and selection process involves the mechanics of calculating results and arriving at a conclusion recommending the acquisition of a DD/DS package.

In this last step of calculation it will become obvious as to reasons why one DD/DS scored higher/lower than others. One of the advantages of using the weighted evaluation technique is that specific weaknesses/strengths can be isolated and analyzed for their effect on the total selection process. Thus, if a package lacked on-line update capabilities its final score could be analyzed and the penalty for lacking this feature could be isolated. Similarly, DD/DS packages can be compared as to their relative scores in particular categories, further highlighting strengths and weaknesses of the packages under consideration.

In final stages of evaluation a selection must be made. At this time it is important to recognize that inspite of all the numbers and weights, the intent to minimize subjectivity does not eliminate the element of individual bias. All of the scores and weights were assigned, not measured. These numbers, by their very nature, are qualitative not quantitative measurements. Thus, in order to distinguish between any two given packages in this selection methodology it is recommended that there be at least a five percent (5%) differential in their scores.

In situations where the selection methodology yields two high scoring packages with less than a five percent differential, the evaluation criteria which are more subjective should be scrutinized. The evaluation criteria which may have caused some bias include the last two categories;

Ease of Use and Resource Utilization
Vendor Support

The selection team should analyze and reevaluate the scores assigned in these categories taking into account that these characteristics are very subjective. The following are the categories that should be considered:

Ease of Learning [7.1] - Each DD/DS package is to be evaluated for the relative ease of users to grasp and retain the instructions for using the system. It is difficult, however, to ascertain this qualitative aspect of the DD/DS without experience with the package. It is therefore recommended that the select team consult with existing users of the package and rely on the experience of others.

Level of Expertise Required [7.2] - This evaluation criteria focuses on the number of people required to support the DD/DS, both administratively and technically. The issue of not having experience with the package arises again. Here also this could be solved by consulting other users. However, one must also contend with varying complexity of the data base environment in which the DD/DS is implemented.

Compatability with Existing Procedures [7.3] - Consideration is to be given, in this criteria, to the ability of DD/DS to adapt to particular characteristics of the installation's data processing environment. This might include the ability to supplement or replace existing forms of documentation, compatability with systems development methodologies and any unnecessary constraints the package may impose on naming conventions or other areas of documentation. The problem with this criteria is that it is vague in its scope. The selection team should enumerate each effected area and create sub-headings for each one to further refine the scoring in an attempt to achieve greater objectivity in the evaluation process.

Vendor Stability [8.1] - It is important to ascertain the ability of the vendor to maintain itself as a going concern. The DD/DS plays an important role in the data base environment and users want a vendor who will not fold and leave users to fend for themselves. Obviously, there is a qualitative judgment to be made. Nevertheless, when arriving at a score, consideration should be given the following qualitative factors;

size of the company (people)
profitability
number of years in business

Commitment to Package [8.2] - The DD/DS should be the central repository for data description and an important control mechanism for the DBA. Thus the DD/DS must be closely coordinated with other software components (e.g., DBMS and TP) of the environment and should be enhanced to periodically stay current with an advancing technology. This requires a vendor committed to supporting a Research and Development activity. This assessment must be made on the basis of past history and future capability.

Reliability and Quality of Support [8.3] - The vendor of the DD/DS software package must be capable of providing reliable support for the user. Personnel assigned should be knowledgeable and capable of solving user problems. The assessment of this capability will be largely subjective and must be based upon past history and the experience of other users.

User's Group [8.4] - The majority if not all of the major DD/DS vendors organize conferences and meetings for their users.

Documentation [8.5] - The importance of good documentation can not be overemphasized for the DD/DS package. Generally, the longer the package has been commercially available, the better the documentation.

7. Conclusion

Many organizations have recently recognized the need for automated support of the Data Base Administration function. Much of this support has materialized in the form of commercial software packages, generally referred to as Data Dictionary/Directory Systems. The process of evaluating and selecting a DD/DS from amongst these systems must be carried out carefully and with maximum objectivity. The methodology set forth in this paper, based upon a uniform DD/DS selection criteria, provides the basis for such an evaluation and selection procedure.

8. Bibliography

EDP Analyzer, November, 1974.

Lefkowitz, Henry C., Data Dictionary/Directory Systems, QED Information Sciences, February, 1977, $75.00.

Plagman, B., Altshuler, G., A Data Dictionary/Directory System Within the Context of an ICDB, AFIPS Conference Proceedings, Vol. 41, 1972.

Technical Profile of Seven Data Element Dictionary/Directory Systems, NBS Special Publication 500-3, U.S. Department of Commerce, National Bureau of Standards, February, 1977.

Uhrowczik, Data Dictionary/Directory Systems, IBM Systems Journal, Vol. 12, No. 4, 1973.

FIGURE CAPTIONS

Figure 1. Selected Commercially Available DD/DS Packages

Figure 2. DD/DS Make or Buy Decision

Figure 3. DD/DS Selection Methodology

Figure 4. Uniform DD/DS Selection Criteria - Major Categories

Figure 5. DD/DS Package Classification

Figure 6. Assigning Weights to Levels of Criteria

DD/DS PACKAGE	VENDOR
LEXICON	Arthur Anderson & Co. 69 W. Washington St. Chicago, IL 60602 (312) 346-6262
TOTAL DATA DICTIONARY	Cincom Systems, Inc. 2300 Montana Avenue Cincinnati, OH 45211 (513) 662-2300
INTEGRATED DATA DICTIONARY	Cullinane Corp. 20 William Street Wellesley, MA 02181 (617) 237-6601
DB/DC DICTIONARY	IBM Corp. 1133 Westchester Avenue White Plains, NY 10604 (914) 696-1900
CONTROL 2000	MRI Systems Corp. 12675 Research Blvd. Austin, TX 78759
DATAMANAGER	MSP, Inc. 594 Marrett Road Lexington, MA 02173 (617) 861-6130
DATA CATALOGUE	Synergetics Corp. One DeAngelo Drive Bedford, MA 01730 (617) 275-0250
UCC TEN	University Computing Co. 8303 Elmbrook St. Dallas, TX 75247 (214) 688-7100
PRIDE/LOGIK	M. Bryce Assoc. 1248 Springfield Pike Cincinnati, OH 45215 (513) 761-8400

FIGURE 1. Selected Commercially Available DD/DS Packages

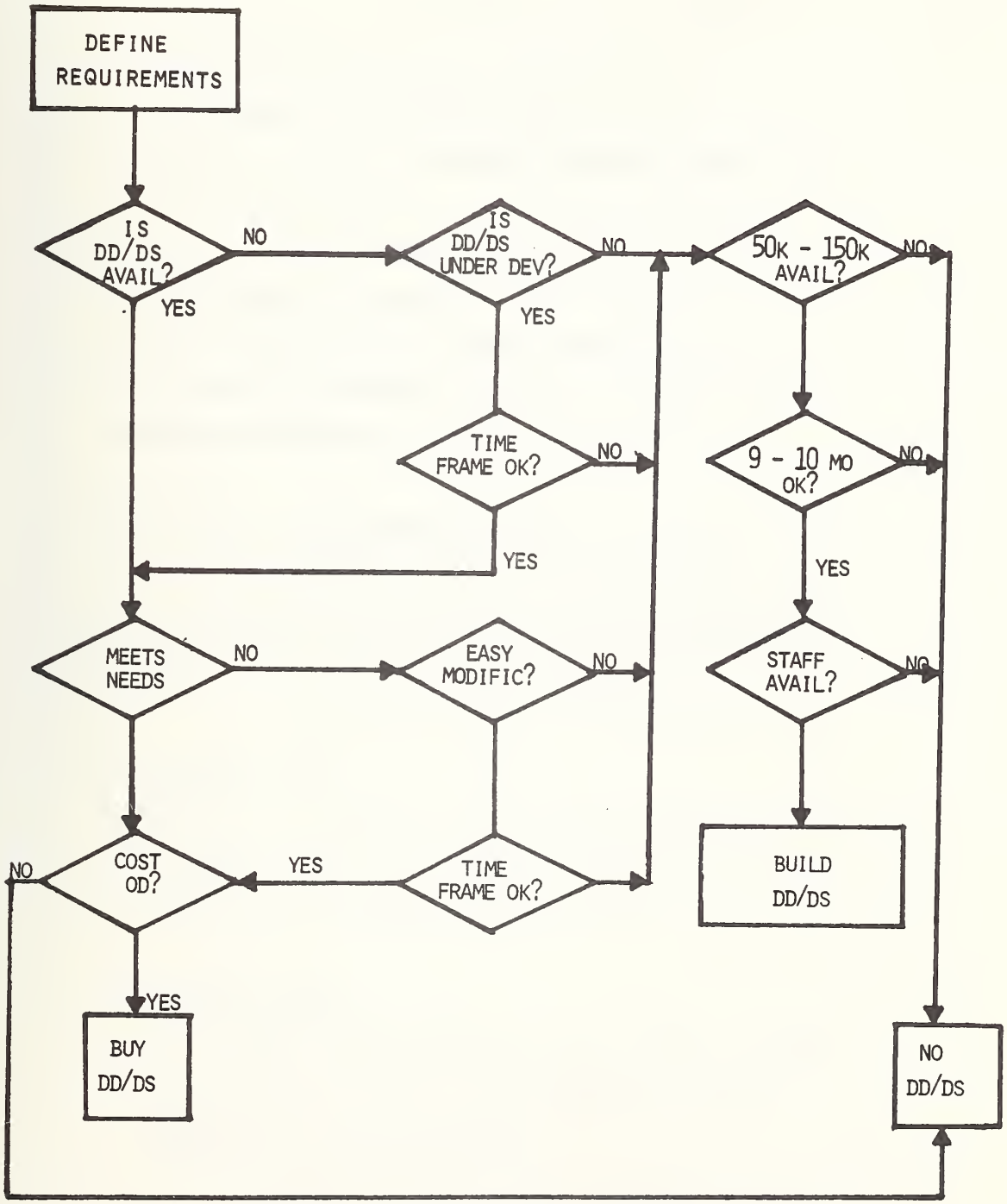


FIGURE 2. DD/DS Make or Buy Decision

1. ESTABLISH SELECTION METHODOLOGY

   - ENUMERATE STEPS
   - SELECT EVALUATION TEAM
   - TRAINING & EDUCATION

2. IDENTIFY AND DESCRIBE SELECTION CRITERIA

   - ENUMERATE CRITERIA
   - DESCRIBE CRITERIA
   - CLASSIFY CRITERIA: MANDATORY/DESIRABLE, ELIMINATING PACKAGES FAILING MANDATORY CRITERIA

3. ASSIGN WEIGHTS TO SELECTION CRITERIA

4. EVALUATE PACKAGES FOR EACH SELECTION CRITERIA

5. CALCULATE SCORES & SELECT DD/DS PACKAGE

FIGURE 3. DD/DS Selection Methodology

1. DATA DESCRIPTION FACILITIES

   THIS CATEGORY INCLUDES CONSIDERATION OF ENTITY DESCRIPTIONS ALLOWED, TYPES OF DATA STRUCTURES SUPPORTED, RANGE OF OPTIONS LEFT TO USER DISCRETION, AND HOST LANGUAGE SUPPORTED.

2. DATA DOCUMENTATION SUPPORT

   THIS CATEGORY COVERS THE RANGE OF ATTRIBUTES PROVIDED TO DESCRIBE DD/DS ENTITIES, AND THE EASE OF EXTRACTING DOCUMENTATION INFORMATION.

3. DATA DEFINITION GENERATION (AUTOMATED)

   A SPECIAL CATEGORY IS ALLOCATED TO EVALUATE THE ABILITY OF THE DD/DS TO GENERATE DATA DEFINITIONS FOR VARIOUS AUTOMATED USES. (e.g., DBMS, HOST LANGUAGE COMPILER, ETC.)

4. SECURITY SUPPORT

   SECURITY IS THE SUBJECT OF THIS CATEGORY, EVALUATED FROM THE VIEWPOINT OF THE ABILITY TO DESCRIBE USER DATA BASE SECURITY, AND FACILITIES TO PROJECT THE DD/DS DATA BASE.

5. INTEGRITY SUPPORT

   THIS CATEGORY ADDRESSES THE ABILITY OF THE DD/DS TO SUPPORT, WITH SPECIFIC FACILITIES, THE DATA BASE ADMINISTRATION RESPONSIBILITY FOR DATA INTEGRITY.

FIGURE 4. Uniform DD/DS Selection Criteria - Major Categories

6. USER INTERFACES/OUTPUTS

CAREFUL CONSIDERATION IS GIVEN IN THIS CATEGORY TO THE QUALITY AND COMPLETENESS OF STANDARD REPORTS, PROVISIONS FOR AD HOC REQUESTS AND ON-LINE FACILITIES OF THE DD/DS.

7. EASE OF USE & RESOURCES UTILIZATION

A SEPARATE CATEGORY IS DEVOTED TO THE EASE OF USE OF THE DD/DS BOTH FROM A PERSONNEL AND AN OPERATIONAL STANDPOINT.

8. VENDOR SUPPORT

ALL AREAS OF VENDOR SUPPORT ARE ADDRESSED IN THIS CATEGORY.

FIGURE 4. (contd)

DD/DS TYPES	INDEPENDENT SYSTEMS	DEPENDENT SYSTEMS
PASSIVE SYSTEMS	LOGIK (PRIDE)	DBD EASTERN AIRLINES
INTEGRATED SYSTEMS	LEXICON DATA CATALOGUE 2 DATAMANAGER	INTEGRATED DD DB/DC DICTIONARY UCC-10 TOTAL DICTIONARY CONTROL 2000

PASSIVE - NO DBMS INTERFACES

INTEGRATED - DBMS DDL GENERATION (AT LEAST)

INDEPENDENT - NOT DEPENDENT ON A DBMS FOR DATA BASE MANAGEMENT

DEPENDENT - DEPENDENT ON A DBMS FOR DATA BASE MANAGEMENT

FIGURE 5. DD/DS Package Classification

CRITERIA	WEIGHTS - LEVELS					SCORES					NOTE
	1	2	3	4	5	(1)	(2)	(3)			
4. SECURITY SUPPORT	10										
4.1 Security Attributes of DB		32									
4.1.1 Included in DBMS Interface			42			0	0	10			
4.1.2 Independent Security Support			58			3	7	0			
4.2 Security Support (DD/DS)		68									
4.2.1 Password Control			62								
4.2.1.1 Global				58		5	0	10			
4.2.1.2 Data/System Entities				42		5	2	10			
4.2.2 Functional Access Control			38			5	0	10			

FIGURE 6. Assigning Weights to Levels of Criteria

Copyright 1975
DBD Systems, Inc.

Some History of Traditional Trade Documents

Robert J. Porter

Eastman Kodak Company
Rochester, New York

Most documents and procedures used in modern domestic and international trade date back hundreds and thousands of years. Trade and commerce have prospered and increased despite increasing difficulties. The development of CARDIS (Cargo Data Interchange Systems) standards with direct aim on timely electronic interchange is of enormous importance today.

Key words: Trade documents; trade data; bill of lading; general average; inventory records; commercial invoice; Shipper's Export Declaration; drafts; letters of credit; data; computerized international documentation system; CARDIS.

Good morning, ladies and gentlemen.

It is indeed a pleasure to be here to see first hand, and for me the first time, the facilities and operations of this prestigious and effective organization.

The greatest kick I get out of this occasion, though, is to realize that after many thousands of years we are on the verge of establishing and publishing world standards for the exchange and interchange of trade data.

My first and foremost job here could be very easy. Mrs. McEwen and my good friend and NCITD colleague, George Begnal asked me to present the NCITD slide-tape show: "Your Link to Trade Profits". As you will see, this takes about sixteen minutes of "no work".

In these few minutes, then, I am going to remind you of the origins of some of the traditional documents which men have used with little change to move goods domestically and internationally since the dawn of human history. Please keep in mind that these ideas and documents were devised for animal caravans and primitive short voyages across rivers, lakes and inland seas when five-ton vessels were considered enormous.

At the same time, though, please keep in mind that our ancestors were as smart or smarter than we are, and while the documents and systems I am about to describe are ancient and have developed like Topsy -- with them one can ship anything from anyplace to anywhere, can protect his interest in the cargo, and can be paid for it. I keep reminding myself that <u>any</u> new system must be at least that good.

Perhaps the oldest commercial document -- no one knows its origins -- was the Bill of Lading or Waybill. Besides being a record of carriage, it is also a contract of carriage, one of the title instruments, and, from the far off past, the strongest evidence of the liability and limits of liability of the carrier. This accounts for the fine print and the signatures. In a traditional, but modern, international sale today it is not uncommon to need three or four bills of lading for a single transaction -- each requiring different classifications of the same goods and each probably denoting different limits of liability and other terms.

To some people it may be a slight surprise to learn that something now considered by most to be included in marine insurance, "General Average" terms are also so old that the origin is really not known.

In the very earliest development of water trade, cargo owners often physically accompanied their goods on the voyage.

Before some genius thought of "General Average", many vessels sank with all goods and hands because no merchants would allow his cargo to be jettisoned by the captain in order to try to save the vessel and people in heavy seas.

The terms of General Average are very simple -- to ancient people. The captain owned and risked his vessel while each merchant owned and risked his cargo. Under the terms of general average, in the event of loss or damage due to a captains' need to react to natural force, each party shares to reimburse the loser or losers -- including the captain.

Inventory records and some form of commercial invoice are so shrouded in history that we don't know how early they appeared.

We do know that about 2,500 years ago kings, princes, and others collected "taxes" on exports and imports -- and sometimes tried to control them as well -- leading to what we know as the Shipper's Export Declarations, various kinds of export licenses, and, in some countries, import licenses.

Demand and documentary drafts were early developments some two thousand years ago -- with letters of credit following right along.

Very early attempts were made by ruling people, some successful and some not, to collect data as well as "taxes". This led to the "classification" of goods and various "Customs" operations.

Right now the average U.S. Exporter deals with the following "historic" documents and more.

1. Commercial Invoice
2. Packing List
3. Export License
4. Inland Bill of Lading
5. Shipper's Export Declaration

6. Certificate of Origin
7. Bank Draft - possibly letter of credit
8. Delivery Order
9. Dock Receipt
10. Ocean Bill of Lading - Airway Bill
11. Consular Invoice - often with fees
12. Forwarder's Bill of Charges

No mention has been made of all the documents behind the exporter, such as records of acquisition, manufacture, inventory, packing, sales analysis, accounts receivable, etc. And need it be said that the goods in a single shipment may have to be listed, described, or classified in six or more different ways? Also, of course many of the documents must be assembled and attached to other documents -- with linguistic translations sometimes required.

Despite all I've said, domestic and international trade have prospered and grown with demand. And I, for one, working for a forward-looking company was instrumental in developing an automatic computerized international documentation system of amazing capability, which has lasted over 15 years.

Fortunately it is now being redesigned by others to be even better in every way -- in part, at least, due to the recognition and use of CARDIS standards about which you will hear from Bob Cavanaugh after you see the show.

Data Resource Management

Daniel B. Schneider

Department of Justice
Washington, D.C.

"Data Resource Management" has become a computer industry buzz-word referring to the management of data outside the confines of specific hardware and software.

Task Group 17 began to focus upon this term when it became clear that data element directories are merely tools with which technicians and managers might fashion some end result, and that the end result to be fashioned would dictate the characteristics of the tools. We had a choice: either focus upon the tools in all their myriad sizes, shapes and possible uses; or pick a specific end result and discuss the directory as a tool for fashioning it. We chose the latter.

The end result to which we oriented was "data resource management". We selected it for two reasons - it seemed the broadest, most comprehensive purpose for developing directories, and it suited the personal interests of the active participants in our sub-task group.

As soon as we commenced work we realized that while the term "data resource management" has long been bandied about by many people of stature in data processing, no clear and complete definition of the term had yet been put forward.

So we set for ourselves four tasks, as follows:

(1) Develop a concept of data resource management.

(2) Show how data resource management relates to other forms of management, e.g. systems management and data processing management.

(3) Develop a concept of a program for the realization of data resource management.

(4) Show how data resource directories can be developed and used within programs for data resource management.

Our results are being documented in two publications, which should be printed sometime in calendar year 1978. An "executive guide" brief will discuss the reasons for having data resource management, and will tell top managers what they must do if they want to launch a data resource management program in their organizations.

A FIPS Publication Guidelines book will tell senior technicians, supervisors and middle managers just what they should do - step-by-step - if and when they are directed to implement a data resource management program. It will contain, as one of its four parts, a discussion of data resource directories prepared by our sister sub-task group, 17.A.

Data as a Resource

A threshold problem we felt compelled to address was whether data is really a resource, in the conventional sense of "resource". Although data shares with manpower, money, machinery and material the basic attributes of economic cost and usefulness to production, it differs from them all in that it is not subject to normal rules and procedures for allocation once it has been acquired.

Rather than assert that data is a resource, we elected simply to state it as a premise. Given that premise, we then examined the other established "resources", identified the common traits in their management, and applied these traits to the resource called data.

Thus we conceive the management of the data resource to be parallel to the management of the manpower, money, machinery and material resources. It is a staff activity, apart from such support services as records operations or data processing. Its operating characteristics are management policy, review and control procedures, and record-keeping. The record-keeping refers, of course, to records about the data resource. This is where the data resource directory comes into the picture.

Finally, we take cognizance of the need for data cost accounting, and the recommendations for this by the Commission on Federal Paperwork. Nevertheless, we feel that a program for data resource management is not in the bailiwick of the accountants, but that such a program may facilitate data cost accounting if and when top management decides to pursue that objective.

Our view of data resource management is unique. No published literature supports it. In fact, in those cases where the phrase "data resource management" appears in printed articles, it does so in contexts very different from ours.

But our view makes sense in light of our premise and the parallels to other resources. The benefits we hold out are real, though we candidly admit the difficulty in forecasting and measuring them. We feel that we are uttering the first words about something that is important and that will last. It remains to those who follow us to refine our concepts in light of their experience.

The Need for Standards in the Health Care Field and A Sampling of Current Applications

Sheila M. Smythe
Executive Vice President
Blue Cross and Blue Shield of Greater New York

This particular panel discussion is, indeed, timely. If your hometown newspapers have carried, as mine has, the announcement of HEW recommended national health planning guidelines, you found them replete with references to resource standards, specifically intended to provide a basis for the identification of inefficient use of scarce resources within the health sector.

In summary, these Federal government recommendations deal primarily with maximum hospital bed population ratios; obstetric, pediatric and neonatal intensive care service needs, open-heart surgery, cardiac catherization units, radiation therapy and CAT scanner distribution, and end-state renal treatment centers. The goals of such recommendations are worthy -- the elimination of some 100,000 excess hospital beds over the next seven years, the consolidation of expensive equipment, and the regionalization of sophisticated technological resources, both human and machine. These proposed standards are, however, limited to general hospitals and do not include federal government hospitals, nursing homes, psychiatric facilities long term or chronic care institutions. These standards are reported to be the first of a series dealing with both the quantity and the quality of health care.

Let us pause for a moment and look at some of the meanings given to the word "standard" by Webster's Collegiate Dictionary.

-- Something established by custom or general consent as a model.

-- Something that is well established by usage or widely recognized as acceptable.

-- Something set up or established as a rule for the measure of quantity, weight, extent, value or quality.

-- The basis of value.

-- A rallying point.

Certainly, within the halls of NBS or within ANSI this tends to be a broader range of meanings than is typically applied all at once to a specific area of application. But it is part of the problem in health, for one must, of necessity, cope with a less tagible entity. Health is neither totally predictable, nor is it a necessarily exact science or art toward which precision can always be directed effectively. Part of the problem is also the solution -- human beings. Another part is the knee jerk reaction to legislative mandate.

Disease does not always proceed in accordance with the rules; neither does our willingness to care for our own personal health.

Let's go back to that middle definition of "standard" -- something set up or established as a rule for the measure of quantity, extent, value or quality. What we see this week is an effort to address through the side doors of quantity and extent, the front door issues of value and quality; for the simple reason that quantity and extent are more easily measured.

The problem of how many services, of what kind, by whom, for whom, when, and how are less conducive to simple mathematical formulae, computerization and orderly information interchange. But, just as technology has advanced in this decade, so must the discipline of the human mind be forced, coerced, charmed or led to grapple with the less manageable. The right environment is a crucial factor in this problem solving. Certainly, the crisis aspect of that environment is clearly with us -- but the atmosphere and conduit for solutions is less defined. I am not trying to say that standards are the solutions, but greater emphasis by the health and technological and professional leadership in standards would help. For in front of the global solutions must come action plans and maps for how to do it, and it is here that well developed standards can play a key role.

All of us involved in standards know that the resolutions do not come overnight. The health care field needs, to my mind, an umbrella in the voluntary sector whereby public and private experts can come together to develop the standards or, more importantly, serve as a receptacle for the labors of others and channel the results to the various parties involved. We are too fragmented at present. This is where ANSI and NBS can play an important role lending the forum, dignity and respect for quality endeavors for which they have become increasingly recognized. This fits in quite appropriately with the last definition of the word "standard" that I gave you -- a rallying point. So much for the scene setter.

Our first panelist will address for us the enormous strides that have been taken in a health application field that truly represents life or death to each of us -- human blood.

Dr. Eric Brodheim is head of the Operations Research Laboratory at Lindsley Kimball Research Institute of the New York Blood Center.

Since 1974 he has been Chairman of the American Blood Commission Committee for Commonality in Blood Banking Automation. This Committee has developed and field tested a uniform labelling system for all blood products that incorporates machine-readable optial bar codes to facilitate a uniform and economical approach to automation by blood banks of all sizes while retaining compatibility with non-automated operations. Since 1975 he has been Chairman of an American Blood Commission Task Force charged with creating a national blood data center to be the focus of information and professional expertise relating to blood and blood policy issues. It was his committee that created the recently published American Blood Commission Report -- Toward a National Blood Data Center. He will speak to us on the Identification and Standardization of Blood Products.

Our second speaker, Gerald J. Duffy, is a native Chicagoan and since 1971 has been Vice President of Data Processing and Telecommunications Services for the Blue Cross Association in Chicago. He has also served for several years on a legislative commission initiated by the House of Representatives in the State of Illinois to study the question of computer research sharing between state agencies, universities, other educational institutions and municipal governments. Mr. Duffy will hopscotch over a variety of interest areas with which he has, will or might be involved.

Our third speaker, Ward C. Duel, is registered sanitarian and public health expert. Currently he is Assistant Director for Environmental Public and Occupational Health at the American Medical Association. Significantly, he is also a member of the one area of health with which ANSI is involved at the Board level, namely, the Medical Devices Standards Management Board.

Mr. Duel will speak to us about various aspects of his work with the FDA and ANSI.

APPENDIX A

Third National Symposium on
the
Management of Data Elements in Information Processing

Symposium Committee

Mrs. Hazel E. McEwen PROGRAM CHAIRPERSON	National Bureau of Standards, Technology Bldg., Room B261, Wash., D.C. 20234 (301)921-3157
Mr. Walter Schlenker PROGRAM CO-CHAIRPERSON	General Electric Company 1285 Boston Avenue, Bridgeport, Conn. 06880 (203)334-1012 Ext. 1498
Mr. Harry S. White, Jr.	National Bureau of Standards, Technology Bldg., Room B261, Wash., D.C. 20234 (301)921-3157

Session Chairpersons

Mr. Aaron Hochman	Office of Secretary of Defense, DoD LOGDESMO - Room 7S-69, 200 Stovall Street, Hoffman Bldg. #2, Alexandria, Va. 22332 (703)325-9324
Mr. Charles J. Kenny	U.S. Civil Service Commission Bureau of Manpower Information Systems 1900 E Street, N.W. - Room 6439 Wash., D.C. 20415 (202)254-8885
Mr. William H. Kenworthey, Jr.	Office of the Secretary of Defense Room 2A256, The Pentagon, Wash., D.C. 20331 (202)688-7608
Mr. Walter Schlenker	General Electric Company 1285 Boston Avenue, Bridgeport, Conn. 06880 (203)334-1012 Ext. 1498
Ms. Nevaire Serrajian	Department of Energy, Office of the Asst. Secretary for Environment, E-201, Wash., D.C. 20545 (301)353-4658
Miss Sheila M. Smythe	Blue Cross-Blue Shield of Greater New York, 622 Third Avenue, New York, New York 10017 (212)490-5421

Presenters and Authors of Papers

Mr. Ted M. Albert	Department of Energy, Office of the Asst. Secretary for Environment, G-107 Wash., D.C. 20545 (301)353-3311
Dr. Craig Black	4400 Forbes Avenue, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213 (412)662-3243
Eric Brodheim, Sc.D	Community Blood Council, 310 East 67th Str., New York, New York 10021 (212) 794-3258
Mr. Paul Brown	U.S. Civil Service Commission - BMIS 1900 E Str., N.W., Room 6457, Wash., D.C. 20415 (202)632-4251
Mr. Robert J. Cavanaugh	General Electric Company, International Sales Division, 570 Lexington Avenue, New York, New York 10022 (212)750-3737
Ward Duel, MPh	American Medical Association, 535 North Dearborn Str., Chicago, Illinois 60610 (312)751-6532
Mr. Gerald J. Duffy	Blue Cross Association of Chicago, 840 North Lake Shore Drive, Chicago, Illinois 60611 (312)440-6258
Peter J. Finnerty, Esq.	Sea-Land Services, Inc., P.O. Box 900, Edison, New Jersey 08817 (201)494-2500
Dr. T. Gary Gautier	National Museum of Natural History Smithsonian Institution, Wash., D.C. 20560 (202)381-6360 or 5252
Mr. Stephen M. Gershenson	Commission on Federal Paperwork 1111 20th Str., Wash., D.C. 20582 (202)653-5414
Mr. Edward A. Guilbert	Transportation Data Coordinating Committee, 1101 17th Str., N.W. Suite 309, Wash., D.C. 20036 (202)293-5514
Mr. Clarence Hankerson	Federal Aviation Administration 800 Independence Avenue, S.W. Wash., D.C. 20590 (202)426-8602

Mr. Eugene Hemley	National Committee on International Trade Documentation, 30 East 42nd Street, Suite 1406, New York, New York 10017 (212)687-6261
Mr. Forest W. Horton, Jr.	Commission on Federal Paperwork 1111 20th Street, N.W. Wash., D.C. 20582 (202)653-5414
Mr. Richard Kline, Jr.	Federal Energy Administration, Energy Systems - Room 7420, 12th & Pennsylvania Avenue, Wash., D.C. 20461 (202)566-9811
Mr. Robert J. Porter	Eastman Kodak Company, Distribution Center, Rochester, New York 14650 (716)458-1000, Ext. 85625
Mr. Daniel Schneider	U.S. Department of Justice 5843 Potomac Avenue, N.W., Wash., D.C. 20016 (202)376-8770
Mr. Stanwyn G. Shetler	Department of Botany, NHB #166 Smithsonian Institution, Wash., D.C. 20560 (202)381-5801
Miss Sheila M. Smythe	Blue Cross-Blue Shield of Greater New York, 622 Third Avenue, New York, New York 10017 (212)490-5421

FUTURE DATA ELEMENT MANAGEMENT CONFERENCES

TO: Harry S. White, Jr.
Associate Director for ADP Standards
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

FROM: (name, address, and telephone number)

___ Please add my name to your mailing list for announcements and information relating to future data element management conferences.

___ I wish to present a paper on the following subject: _____

_____

_____

___ I would like to participate on a panel or in an open forum on the following subject(s): _____

_____

_____

I suggest the following subjects or subject areas for consideration at the future conferences: _____

_____

_____

Other recommendations: _____

_____

_____

_____

_____

NBS-114A (REV. 11-77)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 78-1446	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Management of Data Elements in Information Processing (Proceedings of the Third National Symposium, 1977 September 28-30)

5. Publication Date
1978 April

6. Performing Organization Code

7. AUTHOR(S) Various
Editor: Hazel E. McEwen

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)*

National Bureau of Standards and the American National Standards Institute Committee X3L8

13. Type of Report & Period Covered
Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

Accelerated technological advances in computers and communications make possible the integration of data systems and the exchange of data among them on an expanding scale. However, the full effect of these advances cannot be realized unless the need for uniform understanding of the common information (data elements) and their expression in data systems is recognized and a means provided to effectively manage this information. The increasing interrelationships among data systems of Federal, State and local governments, and with industry and the public add emphasis and dimension to the need for the improved management of data in information processing.

These Proceedings are for the Third National Symposium on the Management of Data Elements in Information Processing held at the National Bureau of Standards on 1977 September 28-30. In these Proceedings, 27 speakers discuss data element management in the field of health care, energy, paperwork management, data dictionary/ directories, data resource management, trade data standards and museum data.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)* American National Standards; American National Standards Institute; data; data base systems; data elements; data management; data processing; Federal Information Processing Standards; information interchange; information processing; information systems.

18. AVAILABILITY

[X] Unlimited

[ ] For Official Distribution. Do Not Release to NTIS

[ ] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Stock No. SN003-003

[X] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
UNCLASSIFIED	155
20. SECURITY CLASS (THIS PAGE)	**22. Price**
UNCLASSIFIED	$8.00

USCOMM-DC 66035-P78

☆U.S. GOVERNMENT PRINTING OFFICE: 1978-261-238/67